APPLICATION OF TARGETING NUCLEASES TO MODELING PROSTATE CANCER GENE REARRANGEMENTS.

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA

Michael D. Nyquist

IN PARTIAL FULFILLIMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advisors: Scott M. Dehm PhD. and Daniel F. Voytas PhD.

June 2014

AKNOWLEDGEMENTS:

I would like express my gratitude to Dan Voytas and Scott Dehm for their guidance and the opportunity to work in their labs. I could have not asked for a better dynamic duo. I would also like to acknowledge Colby Starker and Yingming Li for their respective roles in running well functioning labs and being extremely helpful and imparting their wisdom. I would like to acknowledge other members of the Dehm lab, Luke Brand and Siu Chiu Chan, for teaching me the tools-of-the-trade. I would also like to acknowledge all the members of Voytas Lab, present and former, particularly Nick Baltes, Yiping Qi, and Michelle Christian who helped me get started in learning zinc finger and TALEN technology.

ABSTRACT

Advanced prostate cancer (PCa) treated with androgen deprivation therapy (ADT) eventually relapses to an ADT-resistant disease referred to as castration resistant PCa (CRPC). Recent integrative analyses of PCa genomes have led to the elucidation of potential subtypes that are revelatory to the development of PCa as well as the mechanisms of resistance to ADT and CRPC progression. These studies have confirmed that alterations in the androgen receptor (AR) signaling axis are central to CRPC progression, and have uncovered complex mechanisms by which AR and other components of the AR signaling axis affect, and are affected by, genomic changes and epigenetic transformations. Among the most frequent alterations in CRPC are direct alterations in the AR gene. These AR gene alterations include AR amplification, point mutations, and more recently AR gene rearrangements leading to expression of truncated, constitutively active AR splice variants that are impervious to ADT. Fortunately, the recent development of transcription activator-like effector nucleases (TALENs) has allowed researchers to tailor the genomes of their model systems more rigorously than ever before.

This dissertation presents studies centered on genome engineered cell lines modeling intragenic-*AR* rearrangements that are associated with the production of androgen receptor splice variants. In the second chapter, two *AR* rearrangements associated with ARv567es expression were recreated in PCa cell line R1-AD1 using targeting nucleases. These engineered cell lines expressed high levels of Arv567es that recapitulated the full length AR transcriptome and drove androgen independent growth.

In chapter three, *AR* rearrangements associated with high AR-V7 expression in CRPC cell lines CWR-R1 and 22Rv1 were induced and gene-corrected, respectively, using targeting nucleases. We found that a deletion contained in *AR* intron 1 of CWR-R1 was not sufficient to induce AR-V7 expression. However, genetic correction of a duplicated *AR* intron 3 region found in 22Rv1 decreased AR-V7 levels. These models were developed to determine the underlying mechanisms of AR variant production as well as provide a novel platform with which to study AR variant DNA binding, transcriptional regulation, and clinically relevant aspects of PCa such as biomarker research and precision medicine.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2:

CHAPTER 3:

LIST OF FIGURES

CHAPTER 1:

CHAPTER 2:

LIST OF PUBLICATIONS

1) Nyquist MD, Li Y, Hwang TH, Manlove LS, Vessella RL, Silverstein KA, Voytas DF and Dehm SM. TALEN-engineered AR gene rearrangements reveal endocrine uncoupling of androgen receptor in prostate cancer. Proc Natl Acad Sci USA. 2013 Oct 22;110(43):17492-7.

2) Nyquist MD, Dehm SM. Interplay between genomic alterations and androgen receptor signaling during prostate cancer development and progression. Hormones and Cancer. 2013 Apr;4(2):61-9.

3) Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ, Voytas DF. Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. PLoS One. 2012;7(9):e45383.

4) Lund TC, Kobs AJ, Kramer A, Nyquist M, Kuroki MT, Osborn J, Lidke DS, Low-Nam ST, Blazar BR, Tolar J. Bone marrow stromal and vascular smooth muscle cells have chemosensory capacity via bitter taste receptor expression. PLoS One. 2013;8(3):e58945.

5) Lund TC, Glass TJ, Somani A, Nair S, Tolar J, Nyquist M, Patrinostro X, Blazar BR. Zebrafish stromal cells have endothelial properties and support hematopoietic cells. Exp Hematol. 2012 Jan;40(1):61-70.e1.

CHAPTER 1: General Overview

1.1 Introduction

Prostate cancer (PCa) is the most frequently diagnosed male cancer and second leading cause of male cancer deaths in the USA (Siegel et al, 2013). The androgen receptor (AR) is a master regulator transcription factor in cells of prostatic lineage, and this master regulator function is maintained in PCa cells (Garraway et al, 2006). The frontline treatment for locally advanced or metastatic PCa involves androgen deprivation therapy (ADT), which is achieved by blocking androgen production (surgical or chemical castration with LHRH agonists) or directly antagonizing the AR (antiandrogens). A major limitation is that castration-resistant or castration-recurrent prostate cancer (CRPC) results within 2–3 years due to alterations in the androgen/AR signaling axis. This limitation has begun to be addressed through mechanistic understanding of the changes to androgen/AR signaling that occur during disease progression. For example, it is now established that AR activity persists in CRPC despite ongoing ADT, and the AR signaling axis remains a viable therapeutic target for treating CRPC patients. New therapies that have been developed to re-target AR ligand binding and extend the lifespan of patients with CRPC include the cytochrome P450 c17 (CYP17) inhibitor abiraterone acetate (de Bono et al, 2011), which inhibits androgen synthesis in testes, adrenals, and tumor tissue as well as the next generation antiandrogen enzalutamide (Scher et al, 2012), which suppresses AR transcriptional activity even under conditions of AR overexpression.

Despite these clinical successes, the development of resistance to abiraterone and enzalutamide, and ongoing disease progression for many patients, remains a major

challenge. A greater appreciation of the complex genomic events involved in the initiation and progression of PCa is expected to yield insights into risk and likely routes of therapy resistance taken by individual cells in heterogeneous tumor populations. This could facilitate the individualization and optimization of treatment regimens, which are cornerstones in the concept of precision medicine. To this end, many studies investigating the PCa genome have identified disease- and/or progression-specific alterations in genes that impact the signaling and regulation of the androgen/AR signaling axis. Importantly, included in this list are a myriad of alterations in the AR gene at Xq11-12, which directly alter the regulatory dynamics and biological function of the AR.

1.2 Genomic aspects of Prostate Cancer

1.2.1 Recurrent Genomic Alterations in Prostate Cancer

Gene fusions between the AR-regulated TMPRSS2 gene promoter and the coding region of ETS− family transcription factors, including ERG and ETV1, are among the most frequent recurrent alterations in PCa. These fusions render ETS transcription factors androgen-responsive and under direct control of the AR signaling axis in PCa cells. Their existence in the earliest precursor lesions indicates that these are initiating drivers of PCa and efforts are being made to translate this knowledge into better diagnostic tools and new targeted therapies (Rubin et al, 2011).

Knowledge of the broader genomic landscape of PCa has been greatly expanded in recent years by integrative whole-genome analyses using algorithms that coordinate datasets derived from studies of genome-wide copy number, exome sequence, mRNA expression, and epigenetic modification. These analyses have provided insights into the

mechanisms by which mutations, epigenetic changes, and genome rearrangements conspire to dysregulate genes and pathways in PCa. For example, a recent integrative study demonstrated that the p53/RB pathway was disrupted in 34 % of primary, hormone-naïve PCa and 74 % of CRPC metastases, the PTEN/PI3K pathway was disrupted in 42 % of primary PCa and 100 % of CRPC metastases, and the RAS/RAF pathway was disrupted in 43 % of primary PCa and 90 % of CRPC metastases (Taylor et al, 2010). Similar integrative studies have found that MYC, WNT, FOXA1, SPOP, MLL2, CHD1, and NCOA2 are frequently affected by copy number alterations or point mutations (Grasso et al, 2012; Friedlander et al, 2012; Berger et al, 2011; Beltran et al, 2012). Moreover, these integrative studies have revealed that AR copy number alterations and point mutations do not occur in primary PCa, but occur in 58% of CRPC metastases (Talor et al, 2010). However, when the broader AR signaling axis was considered, alterations were observed in 56 % of primary PCa and 100 % of CRPC metastases, confirming that this master regulator is among the most frequently altered pathways in PCa.

While these integrative studies have revealed important biological insights into PCa, they have relied on targeted approaches that may not reveal the full spectrum of genomic alterations that could occur during disease development and progression. Indeed, recent whole-genome sequencing of seven primary PCa specimens and matched normal tissues demonstrated that important genomic events in PCa are likely underappreciated due to the limitations of targeted approaches such as CGH and exome sequencing. The most striking finding from PCa whole-genome sequencing was the high prevalence of copy number neutral, or balanced, translocations and inversions among the

median 90 rearrangements per genome. (Berger et al, 2012). Some of these rearrangements had interrupted tumor suppressor genes such as PTEN and p53, which would be expected to activate oncogenic signaling through PI3K/Akt and RB pathways, respectively (Berger et al, 2012). It is also interesting to note that some loci did not appear to undergo copy number loss even though inversions and/or point mutations were commonly observed, whereas other loci were frequently deleted (Grasso et al 2012, Berger et al 2011). This suggests that specific chromatin configurations and/or genome architectures may be required for activation or inactivation of genes harbored within these affected regions.

1.2.2 Role of AR in the Genesis of Rearrangements in the PCa Genome

These integrative and whole-genome sequencing studies have raised an intriguing question: why is there a propensity towards structural alteration in the PCa genome? For TMPRSS2–ERG rearrangement-positive cancers, rearrangement breakpoints appear to be enriched near sites of AR and/or ERG binding with open chromatin histone marks, indicating that active transcription driven by these transcription factors may create chromatin environments that are prone to breakage (Berger et al, 2011). There is also gathering evidence to support the notion that AR binding strongly influences higher-order chromatin structure and three-dimensional organization of the nucleus in PCa cells, which may favor genesis of specific rearrangement events by enhancing proximity of discrete genomic domains (Wang et al, 2007; Mani et al, 2009; Lin et al, 2009).

The best evidence supporting a role for AR in shaping deterministic genomic events in the natural history of PCa progression comes from work investigating the

mechanisms underlying recurrent TMPRSS2–ERG fusions. TMPRSS2 expression is controlled by an AR-driven enhancer element just upstream of the TMPRSS2 promoter (Wang et al, 2007). This TMPRSS2 locus is 3 Mb upstream of ETS family transcription factor ERG on chromosome 21. FISH-based studies with molecular probes targeting regions 5′ to the TMPRSS2 breakpoint and 3′ to the ERG breakpoint demonstrated that these loci are brought within close proximity in fusion-negative LNCaP PCa cells upon addition of the natural AR ligand, DHT (Mani et al, 2009). DHT-induced proximity was also observed for the chromosomes 21 and 7 break fusion junctions of the TMPRSS2–ETV1 gene fusion (Lin et al, 2009). Importantly, exposure of LNCaP or the normal prostate epithelial PrEC cell line to gamma irradiation, which induces genotoxic stress, synergized with DHT to induce genesis of TMPRSS2–ERG and TMPRSS2–ETV1 fusions (Mani et al, 2009; Lin et al, 2009). Induction of these fusion events appeared to require DHT/AR-mediated recruitment of various DNA-directed enzymatic activities to the break fusion junctions, which promoted the formation of DNA double strand breaks at these sites. For example, activation-induced cytidine deaminase, the ORF2 endonuclease encoded by long interspersed nuclear elements (LINE-1), and topoisomerase IIB could all be recruited to these AR binding sites, and all have been shown to be important for generation of the fusion events (Lin et al, 2009, Haffner et al 2010). Importantly, DHT-induced proximity and irradiation-induced DNA damage were also shown to lead to generation of the SLC45A3: ETV1 fusion in LNCaP cells, which is a less-frequent PCa gene fusion leading to AR-mediated overexpression of ETV1 in a subset of PCa (Lin et al, 2009; Tomlins et al, 2007). These data imply that many other less frequently occurring rearrangements between androgen regulated

genes and ETS family transcription factors may arise through this same mechanism (Rubin et al, 2011). Importantly, these ETS family fusions may then play a role in promoting further instability of the PCa genome. For example, ERG overexpression increases DNA damage in PCa lines and ERG binding is enriched at rearrangement breakpoints in primary PCa, implicating ERG as a driver of subsequent genomic rearrangements and tumor progression (Brenner et al, 2011; Berger et al 2011).

Epigenetic marks impart an additional layer of regulatory information on top of the information embedded in the PCa genome. Therefore, it is not surprising that many genomic studies have demonstrated that dysregulated epigenetic control accompanies the progression of PCa. Indeed, CRPC genomes display a general overall hypermethylation of CpG dinucleotides when compared with benign prostate (Friedlander et al, 2012). Intriguingly, DNA methylation and genomic changes appear to coordinately alter the regulation of the cell cycle and testosterone metabolism, providing possible specific mechanisms through which DNA methylation promotes PCa progression (Friedlander et al, 2012).

Changes in the epigenetic regulatory environment can potentially change the physiology of the cancer cell and perhaps change the context in which existing and subsequent genomic events are manifest. One example of this is coordinate binding activity between AR and ERG. ERG represses AR transcriptional activity at loci involved in differentiation by recruiting the polycomb group (PcG) protein EZH2, which gives PCa cells more stem cell-like characteristics (Yu et al, 2010). PcG proteins are responsible for maintaining bivalent repression domains to prevent the expression of

lineage genes in embryonic stem cells (Richly et al, 2011). Another recent study associated shared AR and ERG binding sites with transcriptional programs relating to movement and cell proliferation (Chng et al, 2012). This study further showed that HDAC1, HDAC2, and HDAC3 and EZH2 bound to these AR/ERG binding sites. This may have clinical relevance, as HDAC 1, HDAC 2, and HDAC 3 display high expression levels in a majority of PCa and HDAC2 levels in PCa tissues predict shorter PSA-free survival (Weichert et al, 2008). Mixed-lineage leukemia 2 (MLL2), a H3K4-specific histone methytransferase is mutated in 8.6% of PCas (Grasso et al, 2012). MLL2 is also commonly mutated in other cancers and is involved in the regulation of many signaling pathways including nuclear hormone signaling (Guo et al, 2012). Indeed, MLL2 complex members could be immunoprecipitated with an AR antibody along with ERG and FOXA1 (Grasso et al 2012). FOXA1 is mutated in 4% of PCas and is an important pioneer factor for the AR transcriptional program (Barbieri et al, 2012). Interestingly, FOXA1 is preferentially recruited to binding sites enriched in H3K4me1/2 marks, which in turn influences AR binding to nearby androgen response elements (AREs) (Lupien et al, 2008; Lupien et al, 2009). It is tempting to speculate that alterations in MLL2 activity could lead to global alterations in H3K4me1/2 marks, which could in turn affect FOXA1 and AR binding and downstream regulatory processes in PCa cells.

1.3 Truncated AR Splice Variants

Overall, these genomic studies have demonstrated that the AR can be the instigator and the accomplice in effecting cascading genomic and epigenetic changes via chromatin interacting complex members. Interestingly, a pervasive role for the AR in

driving deterministic events in PCa is further evidenced by the high rate of mutation and gene amplification of the AR gene in CRPC. These AR gene aberrations have been shown to alter AR signaling and regulation as well as facilitate AR transcriptional activity despite administration of targeted therapies designed to inhibit the AR (Saraon et al, 2011; Steinkamp et al, 2009; Gottlieb et al, 2012). For example, AR gene amplification leads to overexpression of AR protein, which sensitizes PCa cells to castrate levels of androgens, and can also elicit inappropriate agonist responses to antiandrogens such as bicalutamide (Chen et al, 2004). Similarly, a myriad of point mutations have been described in the AR, many of which occur in the ligand binding domain (LBD) and broaden the repertoire of potential agonists (Gottlieb et al, 2012; Brooke and Bevan, 2009). More recently, rearrangements that alter AR gene structure and splicing patterns have been described in PCa cell lines and xenografts (Li Y et al, 2011; Li et al, 2012). These AR splicing alterations underlie high expression levels of truncated AR variants that lack the AR LBD, which is the target of AR-centered therapies for PCa (Fig. 1-1). The role of AR splice variants in CRPC has been gaining importance in recent years due to the discovery of their prevalence in CRPC tissues (Guo et al, 2009; Hu et al, 2009; Hornberg et al, 2011), and their association with PCa progression and resistance to AR-targeted therapy (Mostaghel et al, 2011; Hu et al, 2012; Li et al, 2012).

Truncated AR splice variants were initially discovered in CRPC cell lines derived from the CWR22 xenograft model of PCa progression (Guo et al, 2009; Hu et al, 2009; Tepper et al 2002; Dehm et al, 2008). They are produced by aberrant splicing and premature translation termination, which results in synthesis of LBD-truncated, constitutively active forms of AR. To date, well over a dozen discrete AR splice variants

have been described (Dehm and Tindall, 2011). Identification and validation of protein translation of these AR variants in PCa cell lines has traditionally involved cloning the mRNA and designing AR variant sequence-specific siRNAs in order to achieve selective knock-down of that truncated AR variant (Guo et al, 2009; Dehm et al, 2008). This selective knock-down strategy has also facilitated functional characterization in PCa cell lines that naturally express high levels of truncated AR variants, such as the 22Rv1, CWR-R1, and VCaP cell lines. In these models, these strategies have revealed that truncated AR variants support androgen-independent expression of AR target genes, and drive androgen-independent growth in a manner that is resistant to anti-androgens (Guo et al, 2009; Hu et al, 2009; Mostaghel et al, 2011; Li et al, 2012; Sun et al, 2010). These key findings have highlighted the importance of understanding the role and origin of AR variants in clinical disease progression.

1.3.1 AR Variant Splicing Mechanisms:

Recent investigation of the mechanisms underlying the synthesis of truncated AR splice variants in CRPC has highlighted genomic rearrangements as a previously-unrecognized class of alterations affecting the AR gene (Li Y et al, 2011; Li et al 2012). For example, our group reported that the 22Rv1 cell harbored a 32kb intragenic duplication of a segment encompassing AR exon 3, which harbors cryptic exons 2b, CE1, CE2, and CE3 (Fig. 1-1). All of these cryptic exons have been shown to splice with high efficiency downstream of AR exon 3 in this cell line; therefore this rearrangement is an attractive explanation for these altered AR splicing patterns. Interestingly, the CWR22Pc cell line is an androgen-dependent sub-line derived from the same parental CWR22

xenograft as 22Rv1 cells. When CWR22Pc cells were cultured under castrate conditions, rare cells harboring this intragenic duplication emerged and continued to display enrichment over several weeks of culture *in vitro*. Importantly, outgrowth of these duplication-positive cells coincided with increasing expression of truncated AR variant mRNA and protein, indicating that expression of these species was restricted to the cell population harboring this underlying rearrangement in the AR gene (Li Y et al, 2011). These data provided the first clues that alterations in AR splicing may be due to underlying alterations in the structure of the AR gene.

An additional CRPC cell line in which truncated AR variants were identified and functionally characterized is the CWR-R1 cell line. In these cells, expression of the AR-V7, encoded by contiguously-spliced AR exons 1, 2, 3 and CE3, was shown to be enriched in CWR-R1 xenografts grown in castrated vs. intact mice (Guo et al, 2009). We recently developed a strategy for paired-end massively parallel sequencing following hybridization-based enrichment of the AR locus, which revealed a 48kb deletion in AR intron 1 in these cells (Li et al, 2012). Interestingly, quantitative analysis of this deletion indicated that the CWR-R1 cell line is heterogeneous, with a variable proportion of cells harboring this deletion. However, long-term culture under castrate conditions resulted in enrichment for this deletion-positive population to near-uniformity. As with the 22RV1 model, the expression level of the AR-V7 variant in the CWR-R1 cell line was directly proportional to the percentage of cells harboring this deletion (Li et al, 2012). Similarly, derivation of discrete sub-clones from this heterogeneous parental population revealed that high levels of truncated AR variant expression was restricted to the deletion-positive cell population (Li et al, 2012).

In addition to these CRPC cell lines, there is also strong evidence to support a role for AR gene rearrangements in clinical CRPC metastases and xenograft tissues derived from clinical CRPC metastases. For example, deletions and duplications involving the AR gene directly alter gene copy number, and imbalances in copy number have been observed across the length of the AR gene in CRPC, but not primary prostatectomy specimens (Li Y et al, 2011; Li et al, 2012). The LuCaP 86.2 xenograft was derived following surgical resection of a CRPC bladder metastasis (Sun et al, 2012), and this tumor displayed reduction in copy number for AR exons 5-7 compared with flanking exons (Li et al, 2012). High-resolution analysis revealed deletion of an 8.5kb segment encompassing AR exons 5-7 in approximately 50% of the cells within this tumor. Importantly, LuCaP 86.2 is the tumor model in which the truncated AR v567es variant was originally identified (es refers to exon skipping, whereby the splicing machinery skips AR exons 5-7, see Figure 1) (Sun et al, 2010). Although a clear cause-effect has not been demonstrated, it would be a logical to predict that the cells in this CRPC tumor harboring deletion of AR exons 5-7 are those cells expressing the AR v567es variant. The discovery of AR gene rearrangements may also explain why certain tumors favor expression of specific species of truncated AR splice variants. For example, it is tempting to hypothesize that ARv567es expression occurs in tumors with genomic alterations that impair proper splicing of exons 5-7, whereas AR-V7 overexpression arises through genomic rearrangements that promote more efficient splicing of AR exon CE3.

An alternative, non-mutually exclusive, mechanism underlying synthesis of truncated AR splice variants could be acute, ADT-induced, changes in regulation of AR

splicing. Such a mechanism was revealed when blockade of full-length AR expression or activity in late-passage LNCaP cells or VCaP cells resulted in acute increases in truncated AR-V7 protein levels (Hu et al, 2012; Watson et al, 2010). These acute splicing changes were rapidly reversed when expression or activity of full-length AR was restored (Watson et al, 2010). However, although AR-V7 levels were clearly increased under these conditions, total levels relative to full-length AR remained low and these cells remained dependent on full length androgen signaling (Watson et al, 2010). Nevertheless, given the importance of truncated AR variants in CRPC, it will be important to elucidate the mechanisms that regulate this plasticity in AR splicing to understand the interplay between these mechanisms and AR gene rearrangements in CRPC.

1.3.2 Clinical Significance of AR Variants

An important role for truncated AR variants in CRPC progression is supported by several studies demonstrating that metastatic CRPC displays high mRNA and protein expression of truncated AR variants compared with hormone-naïve prostatectomy specimens (Guo et al, 2009; Hu et al, 2009; Hornberg et al, 2011; Zhang et al, 2011). Importantly, western blots of CRPC bone metastases demonstrated that a sub-group of CRPC tumors expressed nearly equivalent levels of full-length and truncated AR variants (Hornberg et al, 2011). Moreover, these tissues displayed a higher AR immunostaining score, a higher proliferative index, and were associated with shorter survival after surgery (2 months vs. 8 months) compared with CRPC tissues that expressed predominantly full-length AR (Hornberg et al, 2011). This finding is in-line with another study reporting

that prostatectomy specimens displaying above-median AR-V7 expression had a decreased probability of PSA progression-free survival (Hu et al, 2009).

The expression of truncated AR variants in a subset of CRPC may be especially relevant in light of new AR-targeted therapies that have recently been approved for patients with CRPC. Abiraterone and enzalutamide both increase overall survival of patients with CRPC, but resistance that exists *de novo*, or resistance that develops during therapy is a major limitation for many patients. In mouse xenograft models, abiraterone slowed disease progression, but expression of the truncated AR v567es variant was associated with development of resistance (Mostaghel et al, 2011). Similarly, enzalutamide resistance is driven by truncated AR variants in the 22Rv1 and CWR-R1 cell lines, both of which harbor underlying AR gene rearrangements (Li et al, 2012), and treatment of late-passage LNCaP cells or VCaP cells with enzalutamide leads to rapid increases in AR-V7 protein expression (Hu et al, 2012). A complete understanding of the role of truncated AR variants in driving resistance to abiraterone and enzalutamide in patients will require a detailed and integrated analysis of AR gene structure and mRNA/protein expression patterns in tissues that display progression during treatment with these new agents.

1.3.3 Biochemical properties of Truncated AR Variants

Nuclear localization of full-length AR is tightly regulated by ligand-binding, and AR exists predominantly in the cytoplasm in the absence of androgens. Classical studies have demonstrated that AR nuclear import is supported by a bipartite nuclear localization signal located in the hinge domain encoded by AR exons 3 and 4 (Zhou et al, 1994).

Therefore, truncated AR variants that do not harbor exon 4-derived sequence would be predicted to remain in the cytoplasm. However, loss of the AR LBD also eliminates the AR nuclear export signal. Recently, it has been shown that the net effect of losing both the canonical AR nuclear import and export signals results in constitutive nuclear localization of truncated AR variants in the absence of ligand (Chan et al, 2012). In addition, this constitutive nuclear localization of truncated AR variants does not require full-length AR. However, it has been demonstrated that the AR v567es variant can facilitate ligand-independent nuclear localization of full-length AR through unclear mechanisms (Sun et al, 2010).

Ectopic expression of truncated AR variants in the androgen dependent LNCaP cell line drives androgen independent AR target expression (Dehm et al, 2008; Sun et al, 2010; Chan et al, 2012) and growth under androgen depleted conditions (Guo et al, 2009; Sun et al, 2010; Watson et al, 2010; Chan et al, 2012). Moreover, androgen-independent growth and AR target gene expression in AR splice variant-driven 22Rv1 and CWR-R1 cell lines are blocked when truncated AR variants are selectively knocked-down. These data indicate that truncated AR variants drive androgen-independent growth by supporting ongoing AR transcriptional activity in CRPC cells. However, it has also been shown that truncated AR variants may have unique target gene specificities, perhaps due to the influence of variant-derived protein sequence located immediately COOH-terminal to the AR DBD. For example, unique targets have been ascribed to AR-V7, including AKT1 (Guo et al, 2009) as well as a set of M-phase genes including UBE2C, which is associated with CRPC progression (Hu et al, 2012). However, some of these apparent differences may be due to differences in AR signaling output as opposed to true

differences in target gene specificity. Indeed, UBE2C and other M-phase-associated target genes appear to be transcriptional targets of both full-length AR as well as truncated AR variants, but these genes display a biphasic pattern with induction at proliferative levels of AR signaling and repression at higher, anti-proliferative levels of AR signaling (Li et al, 2012). More recently, gene expression profiling experiments in CWR-R1 cells were designed to compare genes responsive to truncated AR variants or proliferative doses of DHT. In this set of experiments, the gene expression profile supported by truncated AR variants appeared to be a subset of the broader androgen/AR transcriptional program (Li et al, 2012). In future experiments, it will be important to elucidate the sets of genes that are regulated by truncated AR variants in clinical tissues, as this may reveal new therapeutic targets downstream of the AR signaling axis.

1.4 Exploiting Genome Alterations to Define PCa Subtypes

A major goal of cancer genomics is to understand the genomic changes that underlie the development and progression of PCa. This is expected to improve prognostic and predictive accuracy for PCa patients. In particular, classification of clinically-relevant subtypes is likely to prove invaluable in a clinical setting by informing treatment decisions. However, the exact criteria of what constitutes a subtype may be difficult to define. Separating tumors based on mutually exclusive genomic aberrations or a coordinated set of aberrations that accompany distinguishing physiologies and also have bearing on prognostic outcomes and treatment decisions would be ideal. However, even if tumors could be categorized into informative subgroups, PCa is a heterogeneous disease and it is possible that more than one subgroup may be present in a given

individual. This is especially relevant in light of the observation that multifocal PCas often have different genomic rearrangements occurring between foci (Mehra et al, 2007). Additionally, metastatic PCa also appears to have clonal subgroups that respond differently to treatments (Ruiz et al, 2011). Even more daunting is the possibility that mutually exclusive genetic events defined as subtypes in the development of PCa may not inform treatment decisions or have bearing on prognosis. Furthermore, genetic events that occur as an adaptive response to ADT may be the eminent factors to consider when choosing a therapy; this suggests it may be helpful to group cancers based on subtypes that result from acquired genomic changes causative of ADT resistance.

1.4.1 PCa Developmental Subtypes

TMPRSS2-ERG fusions occur in roughly half of all PCa. Therefore, one obvious PCa subtype is TMPRSS-ERG+/- or, more generally, the presence or absence of any AR-regulated or otherwise highly-active gene promoter fused to any Ets family protein (Rubin et al, 2011). Recently, genomic studies have provided insights to mutations that appear to be important drivers in Ets- tumors. For example, SPOP mutations appear to be restricted to primary PCa and CRPC that do not harbor TMPRSS2-ERG fusions (Grasso et al, 2012; Barbieri et al, 2012). Similarly, copy number loss or mutation of CHD1 was found to occur in 5.2% of tumors but most frequently in Ets- tumors (Grasso et al, 2012). Of note, CHD1 binds to H3K4me histone marks and is commonly deleted in PCa contributing to an invasive cell phenotype (Huang et al, 2012). ETS2, a putative tumor suppressor, is also deleted in tumors with the deletion of the 3Mb region between

TMPRSS2 and ERG; however, this locus is also mutated at some frequency in TMPRSS2-ERG fusion-negative cancers (Grasso et al, 2012).

SPINK1 overexpression is another PCa alteration that appears to be mutually exclusive with Ets+ cancers (Tomlins et al, 2008) and constitutes a subtype that occurs in 10% of all PCas. Notably, SPINK1 shares sequence homology with EGF suggesting that it could act as an EGFR agonist. Moreover, overexpression studies have demonstrated that SPINK1 mediates increased invasiveness in vitro (Tomlins et al, 2008), and knock-down of SPINK1 expression with shRNA or antibodies that neutralize SPINK1 have been shown to suppress cellular invasiveness (Ateeq et al, 2011). Additionally, EGFR and SPINK1 suppression has an additive effect in inhibiting cellular invasion in SPINK+/Ets- cell line 22RV1 (Ateeq et al, 2011).

1.4.2 PCa Treatment Subtypes

Genomic changes that occur as a result of ADT may be useful for defining treatment-specific PCa subtypes. Because these would occur later in the natural history of PCa progression, these may be constrained or promoted by preceding genomic alterations that characterize specific PCa developmental subtypes (Fig. 1-2). For instance, it is unclear whether certain truncated AR variants would be able to interact and assemble into a functional complex with the same co-regulator proteins that interact with full-length AR. An example of this is FOXA1, which interacts with full-length AR via the DNA binding domain and hinge region (Gao et al, 2003). Because the AR hinge region is missing from most truncated AR variants (with the exception of ARv567es), it is possible that FOXA1 pioneer activity may not have the same impact on truncated AR

variants as full-length AR. Similarly, the Ets factor ETV1 interacts with the AR ligand binding domain (Shin et al, 2009), which is absent in all truncated AR variants. Interestingly, many of the studies involving ectopic expression of truncated AR variants have been performed in the LNCaP cell line, which displays overexpression of ETV1, including a study which found that full length AR was required for androgen-independent function of truncated AR variants (Watson et al, 2010). In contrast, DAXX, which is involved in cellular differentiation and proliferation, regulates AR activity by binding to the AR NTD (Lin et al, 2004). Therefore, DAXX may be able to interact with and regulate activity of truncated AR variants in a manner similar to full-length AR. Interestingly, SPOP has been shown to bind to and regulate DAXX (Kwon et al, 2006), suggesting a potential interplay between SPOP and the AR in PCa. In addition SPOP has also been shown to regulate SRC-3, an important AR co-regulator overexpressed in CRPC (Li C et al, 2011; Wang et al, 2005).

Changes to the necessity or nature of AR signaling inductive to CPRC such as bypass pathways or changes to the AR itself such as mutations to change ligand specificity, amplifications to sensitize AR to ligand, and rearrangements that induce truncated AR splice variants may occur regardless of developmental subtypes and potentially be the dominant factor with regards to ADT response (Saraon et al, 2009; Pienta et al, 2006). For example, AR amplification can occur regardless of TMPRSS2-ERG fusion or other developmental subtype status (Bismar et al, 2012). However, it may be significant that AR gene amplification is associated with copy number gain of NCOA2 and a higher overall incidence of genomic copy number alterations (Friedlander et al, 2012). AR gene amplification can also increase AR levels and sensitivity to

residual DHT levels and lead to bicalutamide resistance (Chen et al, 2004). Alternatively, inappropriately expressed biosynthetic enzymes that enable intratumoral synthesis of androgens, reviewed extensively elsewhere (Cai et al, 2011), are likely to drive ongoing AR signaling without structural changes to the AR gene locus. In these cases, it is anticipated that tumors harboring AR gene amplifications or altered expression of DHT biosynthetic enzymes would respond to therapy with abiraterone or enzalutamide. Indeed, enzalutamide was developed with the goal of addressing the challenge of treating CRPC driven by AR protein overexpression.

In contrast to amplification- or mutation-based alterations in full-length AR activity, evidence suggest that cancers driven by truncated AR variants may not require full length AR and would thus be resistant to treatments targeting the LBD or androgen synthesis (Barbieri et al 2012; Hu et al, 2012; Li et al, 2012). This highlights the need to develop drugs that target the AR NTD or DBD, which are encoded by the first three exons of the AR gene. One such drug is EPI-001, which has been shown to bind the AR NTD and prevent transcriptional activity of full-length AR by blocking interactions with important AR coactivators (Andersen et al, 2010). Additionally, changes in growth factor signaling axes such as EGFR or IGF1, reviewed elsewhere (Saraon et al, 2011), may activate AR signaling in the absence of ligand or bypass AR signaling altogether by activating mitogenic signaling pathways and render tumors resistant to antiandrogen therapies. Importantly, AR gene amplification events and AR gene rearrangement events may not be mutually exclusive, as AR gene copy number imbalances have been observed in CRPC xenografts and clinical metastases displaying overall increases in AR copy number (Li et al, 2011; Li et al, 2012).

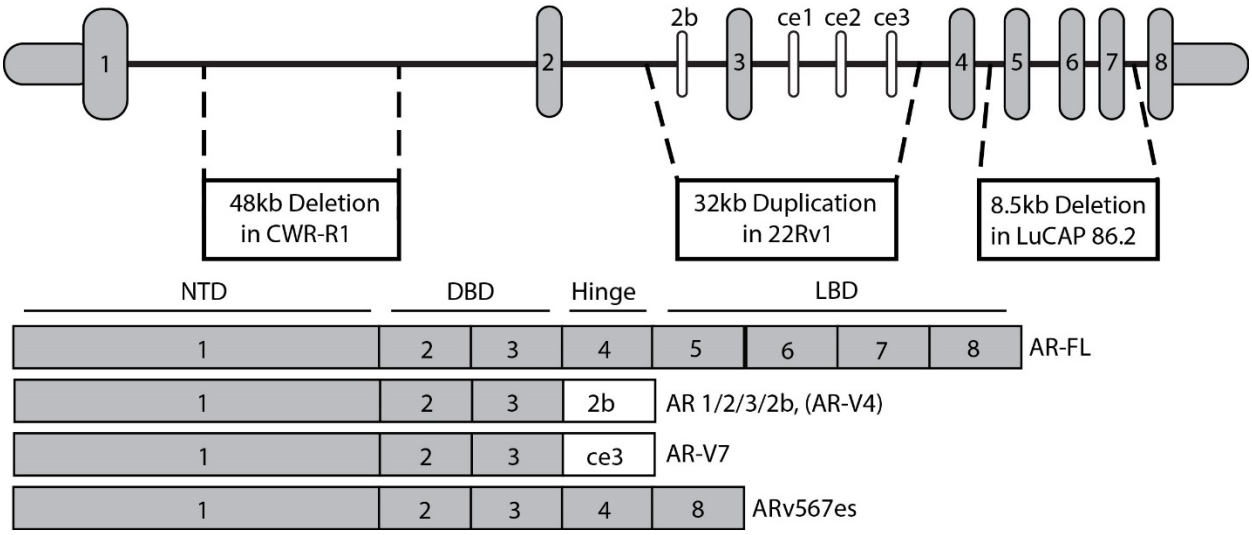
1.5 Future Perspectives

Recent advances in technology and increased sophistication have illuminated complex changes that occur in PCa and possible patterns that can be exploited therapeutically. However, *AR* alterations that drive ADT resistance, such as point mutations, copy number increases, and intragenic rearrangements have not been linked to other contextual alterations in the PCa genome. Whole genome sequencing on CRPC metastases is anticipated to yield insights into the context and patterns of resistance to the expanding repertoire of drug options for ADT.

The complexity of AR regulation necessitates the development of rigorous cell models able to parse the effects of intragenic *AR* rearrangements. Genome editing technology has afforded researchers with the tools to create isogenic models of disease that recapitulate genomic events associated with a phenotype, allowing for greater control of unknown contextual effects. In the work presented here, *AR* is modified to recreate AR-V associated genomic lesions in this fashion.

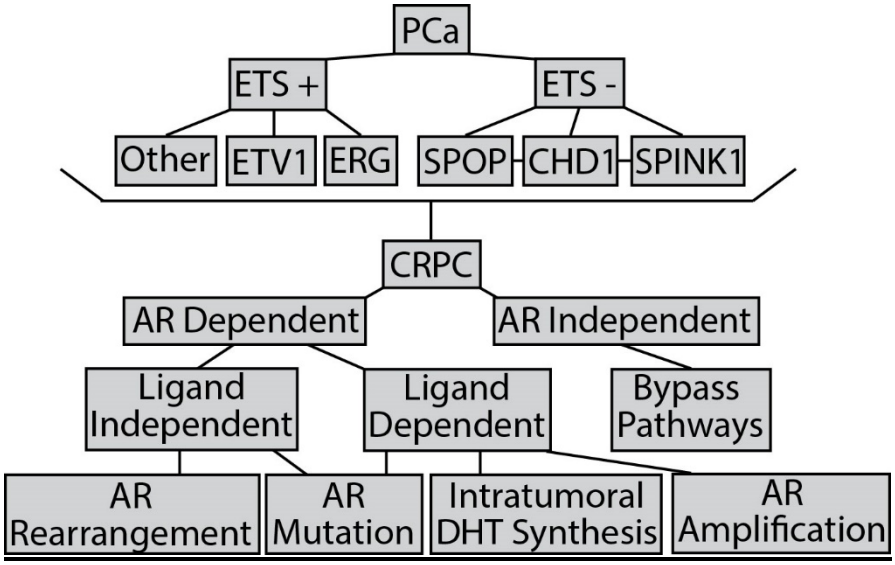
1.6 Figures

Figure 1-1 Genome rearrangements associated with AR variants



A schematic of the AR gene structure is shown with cryptic exons (not to scale). Genomic regions involved in AR gene rearrangements associated with enhanced expression of truncated AR variants are indicated with dashed lines (above). The domain structure of full length AR and truncated AR variants with included exons are indicated (below). Regular AR exons are in gray, cryptic exons in white. NH_2-terminal domain (NTD), DNA-Binding Domain (DBD), Ligand Binding Domain (LBD).

Figure 1-2 Prostate cancer subtypes associated with development and progression



The relationship between PCa developmental subtypes is depicted on top. The relationship between developmental subtypes and treatment subtypes is largely unknown but subsequent changes to AR in response to ADT can be organized into the hierarchical categories depicted on the bottom.

Reprinted with permission from:

Nyquist MD, Dhem SM (2013) Interplay between genomic alterations and androgen receptor signaling during prostate cancer development and progression. *Hormones and Cance*r **4**(2):61-9

CHAPTER 2: TALEN-Engineered AR Gene Rearrangements Reveal Endocrine Uncoupling of Androgen Receptor in Prostate Cancer

Androgen receptor (AR) target genes direct development and survival of the prostate epithelial lineage, including prostate cancer (PCa). Therefore, endocrine therapies that inhibit the AR ligand-binding domain (LBD) are effective in PCa. AR transcriptional reactivation is central to resistance, as evidenced by efficacy of AR retargeting in castration resistant PCa (CRPC) with next-generation endocrine therapies abiraterone and enzalutamide. However, resistance to abiraterone and enzalutamide limits efficacy for most men, and PCa remains the second leading cause of male cancer deaths. Here we show that AR gene rearrangements in CRPC tissues underlie a completely androgen-independent, yet AR-dependent, resistance mechanism. We discovered intragenic AR gene rearrangements in CRPC tissues, which we modeled using transcription activator-like effector nuclease (TALEN) mediated genome engineering. This revealed that these AR gene rearrangements blocked full-length AR synthesis, but promoted expression of truncated AR variant proteins (AR-Vs) lacking the AR LBD. Further, AR-Vs maintained constitutive activity of the AR transcriptional program and a CRPC growth phenotype independent of full-length AR or androgens. These data demonstrate that AR gene rearrangements are a novel resistance mechanism by which AR transcriptional activity can be uncoupled from endocrine regulation in CRPC.

2.1 Introduction

Prostate cancer (PCa) is the most commonly diagnosed cancer in men and is the second leading cause of male cancer mortalities (Siegel et al, 2013). The androgen receptor (AR) is a steroid receptor transcription factor that drives prostate development and homeostasis and is crucial for PCa growth and survival (Garraway and Sellers, 2006). The AR gene is located on the X chromosome, and encodes a modular protein consisting of three major domains: exon 1 encodes the NH_2-terminal domain (NTD), exons 2 and 3 encode a central DNA binding domain (DBD), and exons 4-8 encode the COOH-terminal ligand binding domain (LBD). The NTD is responsible for the majority of AR transcriptional activity, but this activity is suppressed by the LBD unless AR is bound to testosterone or dihydrotestosterone (DHT) (Callewaert et al, 2006; Christiaens et al, 2002; Dehm et al, 2007; He et al, 2004; Jenster et al, 1995). Currently, advanced or metastatic PCa is treated by systemic inhibition of androgen synthesis and antiandrogens that bind to the AR LBD (Ryan et al, 2011). Despite robust responses to these endocrine therapies, castration resistant prostate cancer (CRPC) inevitably develops concurrent with AR transcriptional reactivation (Attard et al, 2011; Chen et al, 2009).

Identification of AR overexpression and high tissue androgen levels as mechanisms driving AR reactivation in a subset of CRPC tumors led to the clinical development and recent approval of abiraterone and enzalutamide as new endocrine targeting therapies for treatment of CRPC (Attard et al, 2008; Chen et al, 2004; de Bono et al, 2011; Montgomery et al, 2008; Scher et al, 2012; Titus et al, 2005; Tran et al, 2009). However, despite the success of these drugs at improving overall survival, primary and secondary resistance is a major limitation for most patients. Point mutations

in the AR LBD have been implicated in resistance to enzalutamide in CRPC cell line models (Balbas et al, 2013). Similarly, increased intratumoral steroidogenesis has been observed in CRPC xenograft models that have developed resistance to abiraterone (Mostaghel et al, 2011). These findings support continued efforts to block the ligand:LBD interaction in order to achieve durable AR inhibition.

An additional occurrence in CRPC is the production of COOH-terminally truncated AR splice variants (AR-Vs). Diverse AR-Vs have been reported, all of which contain the AR NTD and DBD, but lack the AR LBD (Dehm and Tindall, 2011). Functional studies have demonstrated that AR-Vs can function as constitutively nuclear, constitutively active transcription factors (Chan et al, 2012). AR-Vs have been shown to be enriched in CRPC, and track with poor clinical outcomes (Guo et al, 2009; Hornberg et al, 2011l; Hu et al, 2009; Zhang et al, 2011). Recently, AR gene rearrangements were identified in CRPC cell lines that display high-level expression of AR-Vs (Li Y et al 2011; Li et al, 2012). However, AR-Vs are co-expressed with full-length AR in these cell lines, which has made it challenging to pinpoint the precise contributions of AR-Vs to endocrine therapy resistance and a CRPC phenotype. Here, we report intragenic AR gene rearrangements in tissues from clinical CRPC metastases that completely block full-length AR synthesis. We modeled these intragenic rearrangements using TALEN genome engineering (Cermak, et al, 2011), and found that these rearrangements underlie exclusive expression of truncated AR-Vs and uncoupling of the AR transcriptional program from endocrine regulation. Overall, this study represents the first report of a completely androgen-independent, AR-V-dependent resistance mechanism, and establishes the clinical relevance and a functional context of AR-Vs in CRPC.

2.2 Results

2.2.1 AR-V associated intragenic AR rearrangements in clinical CRPC tissue.

LuCaP 86.2 is a PCa xenograft established from a CRPC bladder metastasis resected from a 79 year-old patient (Kumar et al, 2011). This xenograft expresses full-length AR encoded by exons 1-8, but also expresses an AR-V protein thought to be an alternative splicing product arising from skipping of AR exons 5-7 (ARv567es) (Sun et al, 2010). However, since AR is on the X chromosome, discovery of an 8.5kb deletion of AR exons 5-7 in LuCaP 86.2 tissue indicates that ARv567es and full-length AR expression may be mutually exclusive and restricted to distinct cell populations (Li et al, 2012). In line with this, fluorescence in situ hybridization (FISH) and immunofluorescence staining demonstrated that all LuCaP 86.2 cells harbored a single AR gene copy and stained positive with an antibody specific for the AR NTD (Fig. 2-1 *A,B*). However, high-resolution AR copy number analysis using multiplex ligation-dependent probe amplification (MLPA) assay revealed that the AR exon 5-7 segment encoding the AR LBD was deleted in approximately half of these cells (Fig. 2-1 *C*). The relative ratios of deletion-positive and deletion-negative cells were stable during long-term propagation in castrated mice, as were expression ratios of full-length AR and ARv567es (Fig. 2-1 *C,D*). Together, these data demonstrate the existence of at least two stable CRPC subclones in heterogeneous LuCaP 86.2 tissue, one of which harbors an intragenic deletion of AR exons 5-7. To establish the clinical relevance of this finding, we used a PCR-based method and verified heterogeneity for cells with the 8.5kb deletion

breakpoint in both the LuCaP 86.2 xenograft and in CRPC cells procured by laser capture microdissection of archival patient tissue (Fig. 2-2 *A* and Fig. 2-3 *A*-*C*).

Interestingly, a different PCa xenograft established from CRPC abdominal ascites, LuCaP 136, also expresses ARv567es mRNA and protein (Kumar et al, 2011; Sun et al, 2010). Whole exome re-sequencing of LuCaP 136 genomic DNA has been performed (Kumar et al, 2011), but this did not provide an obvious basis for ARv567es expression, and intragenic deletion of AR exons 5-7 was not detected using PCR (Fig. 2-3 *D*). Therefore, we re-sequenced the 183kb AR gene in LuCaP 136 genomic DNA via hybrid capture followed by Illumina-based massively parallel paired-end sequencing. This analysis revealed a copy-neutral 8.7 kb inversion encompassing AR exons 5-7 (Fig. 2-2 *B*, Fig. 2-4 *C*). In contrast to heterogeneous AR expression in LuCaP 86.2 (Fig. 2-2 *C*,*D* and Fig. 2-5), early-passage LuCaP 136 tissue displayed exclusive expression of ARv567es mRNA (Fig. 2-2 *C*) and protein (Fig. 2-2 *D*), which was consistent with very few cells harboring a normal AR allele (Fig. 2-4 *D*). Although there was no archival patient material corresponding to LuCaP 136, discovery of this intragenic AR inversion was made in tissue that had been propagated for only two passages in non-castrate male mice. Later passages of LuCaP 136, which were serially-propagated in non-castrate male mice, displayed coordinate loss of cells with this AR exon 5-7 inversion allele and AR v567es protein expression (Fig. 2-6).

2.2.2 Targeting nucleases recreate ARv567es-associated genome rearrangements.

To study the functional significance of these novel AR gene rearrangements in CRPC tissues, we pursued genome engineering in a monoclonal sub-line isolated from

the CWR-R1 PCa cell line (Gregory et al, 2001), termed R1-AD1 (Li Y et al, 2013). R1-AD1 cells harbor one intact AR gene copy, express full-length AR, and display growth stimulation in response to androgens and growth suppression in response to AR antagonists including enzalutamide (Li Y et al, 2013). We constructed two TALEN pairs (Cermak et al, 2011; Bogdanove and Voytas, 2011; Miller et al, 2011), one pair targeting AR intron 4 (AR-int4) and one pair targeting AR intron 7 (AR-int7) (Fig. 2-7 *A* and Fig. 2-8 *A*). Site-specific double-strand (ds) DNA cleavage by AR-int4 and AR-int7 TALENs was confirmed by assays that measure mutations introduced by TALEN cleavage and imprecise DNA repair (Fig. 2-8 *B,C*). When both AR-int4 and AR-int7 TALENs were co-expressed, deletion or inversion events involving the AR exon 5-7 segment were observed (Fig. 2-8 *D-F*). Using limiting dilution plating and PCR screening, we isolated and expanded two clonal cell lines derived from R1-AD1, one with a deletion (R1-D567) and one with an inversion (R1-I567) of AR exons 5-7 (Fig. 2-7 *B*). AR exon copy number analysis by MLPA demonstrated that R1-AD1 and R1-I567 cells harbored one copy of each AR exon, whereas exons 5-7 had been deleted in R1-D567 cells (Fig. 2-7 *C*). Furthermore, the genome-engineered R1-I567 and R1-D567 cell lines displayed exclusive expression of ARv567es variant mRNA (Figs. 2-7 *D,E* and Fig. 2-9) and protein (Fig. 2-7 *F*). These data confirm that the rearrangements involving the AR exons 5-7 segment that were discovered in CRPC tissues are causative events underlying expression of ARv567es.

2.2.3 ARv567es drives androgen independence in genome engineered cell lines.

Since deletion or inversion of AR exons 5-7 eliminates the AR LBD, cells harboring these rearrangements should have a growth advantage over androgen-dependent PCa cells under conditions where full-length AR activity is inhibited. To test this, we performed competitive growth assays wherein R1-AD1 cells were transfected with AR-int4 and AR-int7 TALENs and then cultured for 20 days in the presence of androgens (FBS or CSS + 1nM DHT), or under conditions of full-length AR inhibition (CSS or CSS + 1uM enzalutamide). Quantitative PCR was used to measure the enrichment of rare cell populations harboring alleles with targeted inversions or deletions relative to all cells. Maintenance under conditions of full-length AR inhibition resulted in the relative enrichment of deletion- and inversion-positive cells. In contrast, under androgen rich conditions, deletion- and inversion-positive cells displayed relative de-enrichment (Fig. 2-10 *A*).

One explanation for these findings could be that deletion or inversion of AR exons 5-7 functionally inactivates AR, causing an AR bypass mechanism of CRPC (Miyamoto et al, 2012). This would be consistent with the notion that full-length AR is required for AR-dependent CRPC progression, even if AR-Vs are expressed (Watson et al, 2010). An alternative explanation could be that ARv567es, which retains the transcriptionally active AR NTD/DBD core (Chan et al, 2012) (Fig. 2-2 *C*) may be able to drive a completely ligand-independent, yet AR-dependent CRPC phenotype. To test this, we performed reporter assays with an AR-responsive MMTV promoter. In parental R1-AD1 cells, MMTV displayed robust androgen induction, which was blocked by knock-down of full-length AR (Fig. 2-10 *B*). Conversely, in R1-D567 and R1-I567 cells, MMTV displayed constitutively high basal activity, but no androgen induction.

However, high basal MMTV activity in R1-I567 and R1-D567 cells was blocked by knock down of ARv567es (Fig. 2-10 *B*). These data demonstrate that ARv567es is a constitutively active transcription factor in R1-I567 and R1-D567 cells.

We next tested whether constitutive ARv567es transcriptional activity could drive androgen-independent growth in R1-I567 and R1-D567 cells. DHT enhanced the growth of parental R1-AD1 cells, which was inhibited by knock-down of full-length AR (Fig. 2-10 *C,D*). Conversely, R1-I567 and R1-D567 displayed robust androgen-independent growth, which was inhibited by knock-down of ARv567es (Fig. 2-10 *C,D*). These data demonstrate that these AR gene rearrangements underlie a shift from a growth profile that is dependent on androgens and full-length AR to a growth profile that is driven by constitutive ARv567es activity.

2.2.4 ARv567es effects a constitutive form of the androgen/AR transcriptional program.

To determine whether ARv567es was functioning globally as a constitutive transcriptional regulator, we performed gene expression microarray analysis of parental R1-AD1 cells and R1-D567 cells. First, we identified genes that were differentially expressed between full-length AR "on" (1nM DHT) and "off" (vehicle control) states in R1-AD1 cells. Next, we identified genes that were differentially expressed between ARv567es "on" (control siRNA) and "off" (siRNA targeting AR exon 1) states in R1-D567 cells. Overall, the majority of genes that displayed differential expression between ARv567es "on" and "off" states in R1-D567 also displayed a similar directional change between the AR "on" and "off" states in parental R1-AD1 cells (Fig. 2-11 *A*). This was

confirmed for several shared target genes (FKBP5, FASN, and LIMA1) by direct qRT-PCR analysis (Fig. 2-12). In line with this, knowledge-based Ingenuity Pathway Analysis (IPA) revealed that the only significant multi-gene networks displaying differential expression in parental R1-AD1 cells or R1-D567 cells had AR as the prominent central hubs (Figs. 2-13). Together, these findings indicate that ARv567es supports constitutive activation of a transcriptional program very similar to the program regulated by full-length AR. To test this more rigorously, we used Gene Set Enrichment Analysis (GSEA) (Subramanian et al, 2005). Genes that were induced or repressed 1.2-fold by DHT in parental R1-AD1 cells were positively or negatively enriched in the R1-D567 gene expression dataset, respectively (Fig. 2-11 B). When we tested the reciprocal relationship, identical results were observed: genes that were induced or repressed 1.2-fold by ARv567es in R1-D567 were positively or negatively enriched in the R1-AD1 gene expression dataset, respectively (Fig. 2-11 *C* and Tables 2-2, 2-3). Based on these data, we conclude that AR gene rearrangements are a CRPC resistance mechanism whereby the AR transcriptional program can be uncoupled from endocrine regulation.

2.3 Discussion

ARv567es as well as other AR-Vs are expressed in CRPC, but the mechanisms underlying this expression, and their functional significance, has been poorly understood (Dehm and Tindall, 2011). In previous studies, we have shown that intragenic AR rearrangements occur in CRPC cell lines where truncated AR-Vs were first discovered, explaining high-level expression of AR-Vs concurrent with full-length AR in these models (Li Y et al, 2011; Li et al 2012). For example, the 22Rv1 cell line harbors a 35kb tandem duplication encompassing AR exon 3, and this intragenic duplication is associated with high-level mRNA and protein expression of the truncated AR-V7 (also referred to as AR3 and AR 1/2/3/CE3) and AR 1/2/3/2b (also referred to as AR-V4 and AR5) variants (Chan et al, 2012; Guo et al, 2009; Hu et al, 2009; Dehm et al, 2008). The CRPC CWR-R1 cell line harbors a population of cells with a 48kb intragenic deletion of AR intron 1, and these deletion-positive cells display high-level expression of AR-V7 and resistance to enzalutamide (Guo et al, 2009; Hu et al, 2009; Li et al, 2012; Li et al, 2013). However, these findings are associative and have not established a clear cause-effect relationship between AR gene rearrangements and functional AR-V expression. The data presented here demonstrate that CRPC LuCaP 86.2 and LuCaP 136 tissues also harbor intragenic rearrangements, associated with expression of the ARv567es variant. However, this study provides an important breakthrough because it establishes a causal relationship between specific genome rearrangements discovered in CPRC tissues and functionally significant AR-V expression. By modeling these specific intragenic AR gene rearrangements using a novel TALEN genome engineering approach, we have demonstrated that these genomic events underlie a true androgen independent phenotype

through a mechanism of switching AR expression from full-length AR to ARv567es. Furthermore, we have established ARv567es as the active protein driving this androgen-independent phenotype by effecting a broad androgen/AR transcriptional program.

A rearrangement-dependent mechanism of AR-V expression in CRPC tissues contrasts with the observation that AR-V7 expression increases acutely in VCaP and late-passage LNCaP cells in response to endocrine targeting therapy (Hu et al, 2012) supporting growth of VCaP cells when full-length AR is inhibited (Liu et al, 2013). This plasticity in AR-V expression could be attributable to a negative feedback loop, wherein transcription of the AR gene is actively repressed by androgen-bound full-length AR via recruitment of lysine specific demethylase 1 (LSD1) (Cai et al, 2011). Based on this mechanism, AR inhibition would lead to transcriptional de-repression and increased expression of both full-length AR as well as AR-V7. However, unlike AR gene rearrangements, this mechanism does not appear to affect relative expression ratios, and AR-V7 remains a minor constituent of overall AR expression (Hu et al, 2012; Liu et al, 2013).

Tumor heterogeneity and subclonal architecture is not apparent when lysates from cell lines or tissues are assessed for mRNA or protein expression. A frequent finding in CRPC cell lines and tissues, including LuCaP 86.2 and LuCaP 136, is concurrent expression of AR-Vs along with full-length AR (Guo et al, 2009; Hornberg et al, 2011; Hu et al, 2009; Zhang et al, 2011; Sun et al, 2010). Modeling this co-expression by transfecting AR-Vs in LNCaP cells led to the conclusions that full-length AR and ARv567es can physically interact (Sun et al, 2010), that ARv567es facilitates nuclear localization and transcriptional activation of full-length AR under castrate conditions

(Sun et al, 2010), that full-length AR is required for ARv567es to drive features of the CRPC phenotype (Sun et al, 2010; Watson et al, 2010), and that antiandrogens such as enzalutamide can inhibit full-length AR and thereby inhibit any effects of AR-Vs (Watson et al, 2010). Our data highlights the importance of understanding the subclonal architecture of CRPC cell lines and tumor tissue when assessing AR-V expression and function. Indeed, there have been conflicting reports about the association of AR-V expression levels with clinical disease progression (Guo et al, 2009; Hornberg et al, 2011; Hu et al, 2009; Zhang et al, 2011; Zhao et al, 2012), and our data suggests that this could be due to the variable degree to which heterogeneous CRPC tumors are enriched for AR-V expressing cells.

The majority of men progressing on abiraterone and enzalutamide display rising serum levels of the AR-regulated PSA gene, indicating these tumors remain AR-driven (de Bono et al, 2011; Scher et al, 2012; Ryan et al, 2012). This has spurred ongoing development of additional endocrine therapies that act on the AR LBD (Balbas et al; Clegg et al, 2012; DeVore and Scott, 2012). Importantly, the AR gene rearrangements discovered and modeled in this study provide a complete genetic block of full-length AR and represent a previously-unanticipated endocrine uncoupling escape mechanism. Identification of AR-Vs as the ultimate functional outcome of this mechanism highlights a need for development of new therapies targeted to the AR-V core, which is composed of the transcriptionally active AR NTD and AR DBD (Anderson et al, 2010; Dehm and Tindall, 2007). In addition, this work highlights an opportunity to develop AR gene rearrangements as stable biomarkers of resistance to endocrine targeting therapies for prostate cancer.

2.4 Methods

Tissues. Genomic DNA samples from the LuCaP series of PCa xenografts and de-identified clinical CRPCa tissue were obtained from the University of Washington Prostate Cancer Biorepository, which was developed and managed by one of the co-authors (R.L.V.) and has been described in previous publications (Zhang et al, 2011; Kumar et al, 2011; Sun et al, 2010).

Cell lines. The LNCaP (#CRL-1740) cell line was obtained from American Type Culture Collection ATCC (Manassas, VA) and CWR-R1 cells were provided by Dr. Elizabeth Wilson (UNC Chapel Hill, NC). The R1-AD1 cell line (CWR-R1, androgen-dependent 1, referred to as "deletion-negative clone 1" in the original publication) (Li et al, 2012) is a subline derived from single-cell cloning of the CWR-R1 cell line. R1-AD1 cells are androgen responsive and contain a structurally normal copy of the AR gene (Li et al, 2012; Li et al, 2013). Cells were maintained in in RPMI 1640 medium supplemented with 10% FBS in 5% CO_2 incubator at $37^{O}C$.

MMTV-luciferase reporter assays. Promoter-reporter assays with an MMTV-Luciferase reporter were performed as previously described (Dehm et al, 2008).

Cell growth assays. Cell growth was monitored by crystal violet staining as previously described (Li Y et al, 2011). For 20-day growth enrichment assays, cells were electroporated with either a CMV-GFP expression vector and pBluescript stuffer vector or equal masses of all four TALENs. Transfected cells were plated on 6-well tissue

culture plates and grown to confluence. Confluent cells were trypsinized and re-seeded on 6-well tissue culture plates in RPMI 1640 medium supplemented with 10% FBS, 10% CSS, 10% CSS + 1uM enzalutamide (Selleck Chemicals), or 10% CSS + 1nM DHT (Sigma) . Cells were cultured for 20 days under these conditions with trypsinization and re-seeding when cells became confluent. At the end of 20 days genomic DNA was harvested for analysis by quantitative PCR.

Western blots. Western blots were performed as previously described (Li Y et al, 2011) using antibodies specific for the AR NTD (N-20, Santa Cruz Biotechnology) and ERK2 (D-2, Santa Cruz Biotechnology).

Generation of R1-D567 and R1-I567 cell lines. To isolate clonal cell lines with targeted intragenic deletion or inversion of the AR exon 5-7 segment, cells were electroporated with TALENs, plated and allowed to recover for 3 days. Transfected cells were transferred to tissue culture dishes at limiting dilution. When colonies became visible (~3wks to 1month), cells were picked, trypsinized, and plated into 96-well plates. Clones were expanded and split into two separate plates, one of which was used for PCR screening. Genomic DNA was extracted using a QuickExtract kit (Catalog#: QE0905T, Epicentre-Illumina) according to the manufacturer's protocol. Genomic DNA was then used for PCR screening with rearrangement-specific primers (Table 2-1). Positive clones were expanded. Clonal purity of R1-D567 and R1-I567 cell lines were determined by secondary PCR screening, and met criteria of positive PCR signals with rearrangement-specific primers and negative PCR signals with wild-type specific primers (Table 2-1).

Multiplex ligation-dependent probe amplification (MLPA). MLPA assays were performed as described previously (Li et al, 2012).

Gene expression microarray analysis. Gene expression microarray analysis with Illumina Human HT-12 V4 Beadchips was performed as described previously (Li et al, 2013). Samples used in analysis were in biological triplicates. R1-D567 cells were electroporated with either siCNTL or siAR-exon1 (Table 2-1). R1-AD1 cells were electroporated with siCNTL. Electroporated cells were plated in RPMI 1640 medium supplemented with 10% CSS. After 48h, cells were switched to serum free RPMI-1640 medium supplemented with 1nM DHT or 0.1% (v/v) ethanol (vehicle control) for 24 hours. Gene lists were derived from genes displaying an expression increase or decrease of $\geq$1.2 fold ($p = \leq 0.05$ with Benjamini-Hochberg FDR correction) in R1-AD1 1nM DHT vs. EtOH and R1-D567 siCNTL vs. siAR-exon1. Heat maps were generated using Cluster 3.0 software (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). Data are available through NCBI's Gene Expression Omnibus (GSE49196).

Gene set enrichment analysis (GSEA). Lists of differentially-expressed genes were analyzed for enrichment in normalized gene expression datasets using GSEA analysis software v2.0.10 (http://www.broadinstitute.org/gsea/index.jsp). Normalized gene expression data was ranked using the Signal2Noise metric and GSEA was performed against 1000 random gene set permutations. All other analysis options were left at the default setting.

Florescent In-Situ Hybridization (FISH). FISH was performed by the Cytogenetics Shared Resources Laboratory at the Masonic Cancer Center. BAC probe RP11-807F19 (ChrX:66714000-66908813) was labeled with Orange - 552 dUTP (Enzo Life Science) using a nick translation kit (Abbott Molecular) as per the manufacturer's protocol. Precipitated probe fragments were dried and resuspended in 50% formamide hybridization buffer and mixed with a Spectrum Green-labeled commercial probe for the centromere of chromosome X (Abbott Molecular). Frozen cryosection slides of LuCaP 86.2 tissue were thawed and pretreated in a sequence of 2xSSC, ethanol, and acetone in preparation for hybridization with the FISH probes. The probe/hybridization buffer mix was applied to pretreated tissue slides. Slides were placed in a Thermobrite (Abbott Molecular) denaturation/hybridization chamber, denatured for 5 minutes at 75°C, then incubated overnight at 37°C. Slides were washed in a 2xSSC solution at 72° for 15 seconds, and counterstained with 4',6-diamidino-2-phenylindole (DAPI) stain. Slides were visualized on an Olympus BX61 microscope workstation (Applied Spectral Imaging, Vista, CA) using DAPI, FITC and Texas Red filter sets. Images were captured using an interferometer-based CCD cooled camera (ASI) and BandView ASI software.

Immunofluorescence (IF). LuCaP 86.2 tissues that had been passaged in intact (s36n9 and s36n8-1; s and n = number of passages in SCID and NSG mice, respectively) or castrated (s7n2-2 and s7n2-3) mice were frozen in optimal temperature cutting (OTC) media and 6 micron sections were cut onto slides. Tissues were fixed by cold acetone for 5 min, dried at room temperature 10 min, then rehydrated in 1X phosphate buffered

saline (PBS) at room temperature for 10 min. Samples were blocked for 1 hour at room temperature with blocking buffer (0.2% Tween-20 and 2% BSA in 1X PBS) containing 5% normal goat serum (Sigma #G 9023). AR N-20 primary antibody (Santa Cruz) was applied overnight at 4^{o}C at a 1:200 dilution in blocking buffer. Cells were washed 3 times for 5 minutes each in a washing buffer containing 0.2% Tween-20 in 1XPBS. The secondary antibody, goat anti-rabbit Alexa Fluor 488 F(ab')$_2$ Fragment of Goat Anti-Rabbit IgG (H+L) Antibody (Life Technologies), was applied for 1 h at room temperature at a 1:500 dilution in blocking buffer. Slides were washed twice for 5 minutes each in washing buffer and then incubated for 15 minutes of 1ug/mL DAPI followed by a final wash of 5 minutes in 1X PBS. Cells were then mounted with Vectashield mounting media (Vector Laboratories) and images were captured with a fluorescent microscope (Nikon Eclipse TS100) equipped with a 20X objective.

Next-generation paired-end re-sequencing of the 183kb AR gene. Genomic DNA from LuCaP 136 xenograft tumor tissue was subjected to hybrid capture with a custom SureSelect bait library (Agilent) and sequenced at 6000X depth with 2X100bp settings on an Illumina HiSeq 2000 as described(Li Y et al, 2012). Briefly, raw sequence data from HiSeq 2000 was de-multiplexed and filtered using CASAVA 1.8. FASTQ formatted reads were inspected using fastQC (Babraham Bioinformatics Institute). Filtered reads were trimmed (from an initial length of 2 x 100 bp to 2 x 85 bp, removing sequence from the 3′ end) to remove low-quality ends using the FASTQ trimmer tool in Galaxy, then mapped to the hg19 build of the reference genome using Burrows-Wheeler Alignment (BWA)(2). Briefly, for BWA alignment, the seed size was 20 (“-l 20”). Up to two

differences within the seed ("-k 2"), up to 4 differences in each read end ("-n 4"), and up to 1 gap opening in the alignment ("-o 1") were allowed. In the alignment pairing phase, the maximal expected insert size ("-a") was set to be 500. Up to 10 million possible mapping locations ("-o 10000000") were also allowed. Output BAM files from BWA were sorted, followed by removal of potential PCR duplicates using Picard tools (http://picard.sourceforge.net/). Discordantly-mapped read pairs and soft-clip reads were collected and used to identify structural variations via LUMPY (https://github.com/arq5x/lumpy-sv) with parameters: -mw 3, -tt 1e-3, back_distance:20, weight:1, min_non_overlap:85, discordant_z:7, back_distance:20, mean:227, stdev:73.

Transient transfections. Cells were transfected with siRNAs and/or plasmid vectors by electroporation as described previously(Dehm et al, 2008). Briefly, 2-3X10^6 cells were re-suspended in 350 uL of media and mixed with 50 uL of TE buffer containing 120pmoles of siRNA and/or 12ug of DNA. Cell/siRNA/DNA mixes were added to a 4mm gap-width cuvette and subjected to 2X 275 volt pulses of 100ms interval/pulse. Cells were allowed to recover 15 minutes at room temperature before plating. Control siRNA was purchased from Dharmacon (siCNTL siGENOME Non-Targeting siRNA #1, Cat# D-001210-01-50). AR-targeted siRNAs had the following sequences: siAR-Exon1, sense: 5'-CAAGGGAGGUUACACCAAA, antisense: 5'-UUUGGUGUAACCUCCCUUG; siAR-Exon7 sense: 5'-GGAACUCGAUCGUAUCAUU, antisense: 5'-AAUGAUACGAUCGAGUUCC.

RT-PCR and quantitative RT-PCR. RNA was isolated and converted to cDNA as described previously(Li Y et al, 2011). cDNA was used for PCR reactions with Exon 4 fwd and Exon 8 rev primers (Table 2-1) using *Taq* DNA polymerase (Qiagen, catalog #201205) according to the manufacturer's protocol. For quantitative RT-PCR assays, cDNA was used for quantitative PCR reactions with primers specific for FKBP5, LIMA1, FASN, and GAPDH (Table 2-1) using a BioRad iCycler and PerfeCTa Sybr Green FastMix (Quanta Biosciences, 95072-250) according to the manufacturer's protocol.

TALEN nuclease construction. Design and construction of expression vectors for TALENs was carried out as previously described (Cermack et al, 2011). Briefly, TALEN pair target sites were chosen so that each TALEN contained 15 repeat variable di-residue (RVD) modules with a 14-15 base pair spacer in between each TALEN target site. The TALEN RVD modules were assembled from a cloning library (Addgene, Golden Gate TALEN and TAL Effector Kit 2.0, #1000000024) and Golden Gate cloning method. RVD modules were assembled into two fusion vectors (FusA and FusB) per TALEN, one containing the first 10 RVD modules and the second containing four RVD modules. The final TALENs were constructed through a single Golden Gate cloning reaction with the FusA and FusB vectors, the last half RVD module, and the mammalian expression vector backbone pC-GoldyTALEN (Addgene:#38143)(Carlsonet al, 2012).

Genomic DNA PCR and Surveyor (Cel 1) nuclease assays. PCR for detection of targeted deletions or inversions in R1-AD1 cells and for Cel 1 assays were performed

using Phusion-HF Polymerase Kit with HF buffer (Catalog#: M0530S, New England Biolabs) according to the manufacturer's protocol. PCR primers are listed in Table S1. For assays with Surveyor (Cel 1) nuclease, genomic PCR products were separated on 1% agarose gels and the bands were excised and purified using a QIAquick Gel Extraction kit (Catalog#: 28706 Qiagen) according to the manufacturer's protocol. A 400ng aliquot of the purified PCR product was digested using Surveyor Mutation Detection Kit (Catalog#:706025 Transgenomic) according to the manufacturer's protocol. Digests were analyzed on 2% agarose gels.

Genomic DNA PCR. Genomic DNA was isolated from tissues and cell lines using a Nucleospin tissue kit (Catalog#: 740952, Machery-Nagel) according to recommended protocol. PCR with LuCaP86.2 and LuCaP 136 genomic DNA was performed using Taq DNA polymerase (Qiagen, catalog #201205) according to the manufacturer's protocol. Primers for PCR detection of the LuCaP 86.2 deletion junction, the LuCaP 86.2 5' inversion junction, the LuCaP 86.2 3' inversion junction, and wild-type AR locus controls are listed in Table S1.

Quantitative genomic PCR. Quantitative PCR with genomic DNA from 20-day growth enrichment assays was carried out using a LightCycler 480 II (Roche) and Lightcycler 480 SYBR Green I Master Mix (Catalog 04707516001, Roche) according to the manufacturer's protocol. Primers used to detect inversion and deletion junctions are listed in Table S1. Threshold cycle of amplification (Ct) values derived using inversion and deletion-specific primers were normalized against Ct values derived from primers

targeting a control region in AR intron 2 (Table 2-1). Data were transformed using the formula 2^{-ddCt} with the relative level of inversions and deletions at Day 0 arbitrarily set to 1. Quantitative PCR on LuCaP136 genomic DNA was performed using qLuCaP 136 F1 and qLuCaP 136 R1 to amplify inversion junctions and qLuCaP F1 and qAR normal R1 was used to specifically amplify wild type AR. All samples were normalized to an unrelated locus in AR intron 1 using primers qAR intron1 F and qAR intron 1R. Quantitative PCR was performed using the Perfecta SYBR Green Fastmix according to recommended protocol (Quanta Biosciences, Catalog#:95071-012) (primers are listed in Table 2-1).

Cloning and sequencing of PCR and RT-PCR products. AR cDNA products were cloned using the TA-TOPO Cloning Kit (Catalog#:45-0641, Invitrogen) according to the manufacturer's protocol. Genomic PCR products were cloned using the CloneJet PCR cloning kit (Catalog#: K1231 Thermo Scientific) according to the manufacturer's protocol. All cloned plasmids were sequenced by Sanger sequencing with a universal T7 primer.

Microarray sample preparation. Cells were harvested using Trizol (Life Technologies, catalog#:15596-026) according to the manufacturer's protocol. Total RNA was then column purified using an RNeasy Mini Kit (Qiagen, Catalog#: 74104) and 2 ug was submitted to the University of Minnesota Biomedical Genomics Center for Illumina Direct Hybridization array analysis (Illumina, San Diego, CA). RNA quality control was performed using a NanoDrop 8000 spectrophotometer (Thermo Fisher

Scientific,Waltham,MA) and Caliper LabChip GX (Caliper Life Sciences,Hopkinton,MA). Total RNA was converted to amplified biotinylated, antisense cRNA using the Illumina TotalPrep-96 RNA Amplification Kit (Life Technologies, Carlsbad, CA), and 150 ng of biotin-labeled cRNA was hybridized onto Illumina HumanHT-12 v4 Expression Beadchips (Illumina, San Diego, CA) using the HumanHT-12 v4 Expression BeadChip Kit (Illumina, San Diego, CA). Hybridized Beadchips were scanned with an Illumina iScan. Raw intensity data was extracted from iScan scan image files (Illumina) using GenomeStudio software (Illumina), $\log_2$ transformed, and then imported to Partek Genomics Suite 6.6 (Partek Inc., St. Louis, MO). Using Partek GS, raw intensity data was quantile normalized and differential gene expression was determined by one-way analysis of variance (ANOVA) using default settings within the Gene Expression Workflow.

Ingenuity Pathways Analysis (IPA). Lists of differentially-expressed genes were analyzed using Ingenuity Pathway Analysis (IPA, version 8.5, Ingenuity Systems, http://www.ingenuity.com, Core 8.5, Ingenuity Analysis), a commercial application that infers the relationship between the gene set and known pathways. We used default settings, wherein 'Ingenuity Knowledge Base (Genes Only)' was the reference set, and both direct and indirect relationships were considered to identify sub-networks enriched with the selected genes.

2.5 Figures

Figure 2-1. LuCaP 86.2 tissue is heterogeneous for cells with an intact AR gene and cells harboring deletion of AR exons 5-7.

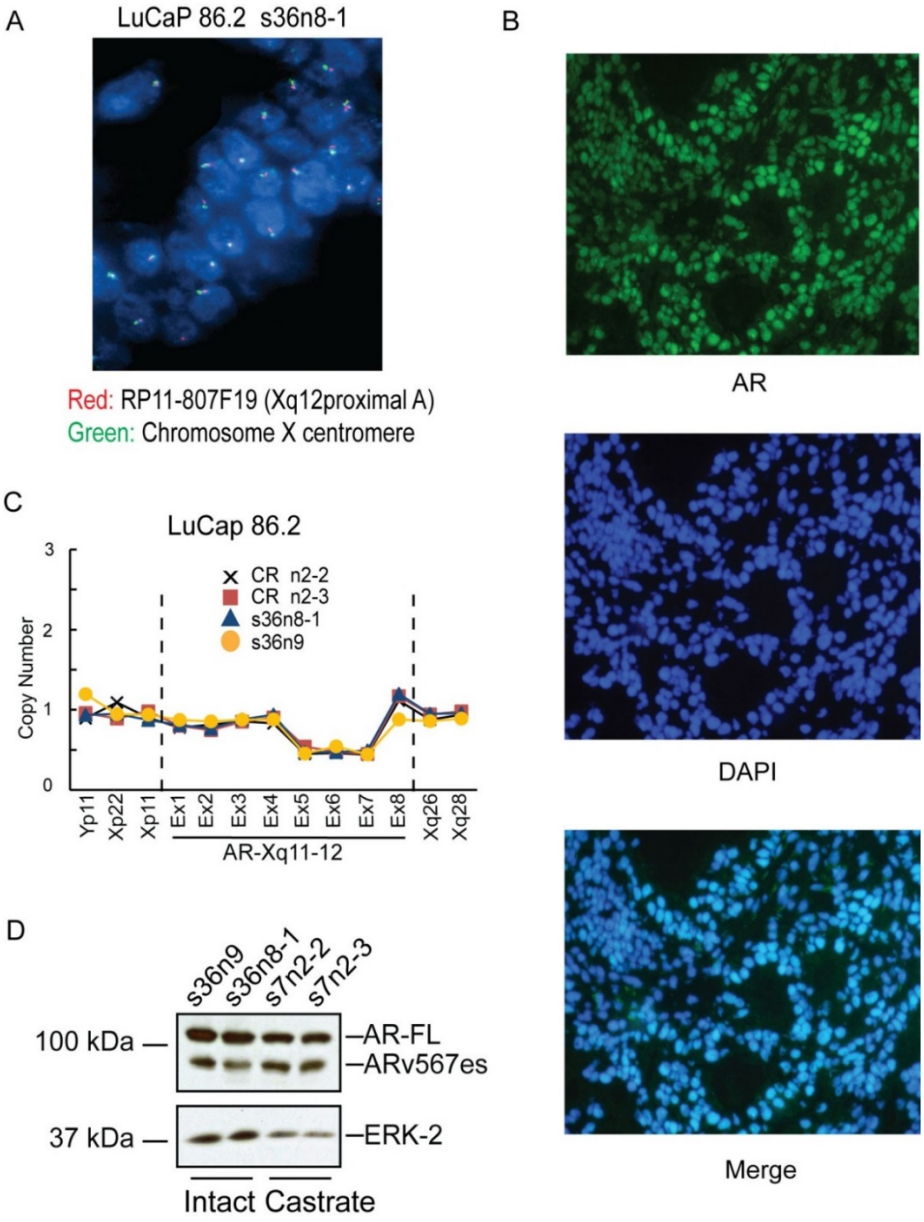


Figure 2-1. (A) Representative image of LuCaP 86.2 tissue subjected to FISH with a probe specific for the AR gene (Spectrum Red signal) and a probe specific for the X centromere (Spectrum Green signal). Evaluation of 200 cells revealed that 193 cells (96.5%) had one red signal and one green signal, 3 (1.5%) cells had one red signal and no green signal, and four cells (2%) had no red signal and one green signal. (B) Immunofluorescence was performed using an antibody specific for the AR NTD (Alexa Fluor 488 signal) and nuclei were stained with DAPI. Representative fluorescence microscopy images are shown for three color channels (Alexa Fluor 488, DAPI, and merged Alexa Fluor 488/DAPI) (C) MLPA genomic copy number analysis in LuCaP 86.2 tissues that had been propagated in intact (s36n9 and s36n8-1; s and n = number of passages in SCID and NSG mice, respectively) or castrated (s7n2-2 and s7n2-3) mice. (D) Western blot for the AR NTD or ERK-2 (loading control) in LuCaP 86.2 tissue propagated in intact or castrated mice. Contributions: (A,D) Oseth LA, Betsy Hirsh B (D) Li Y.

Figure 2-2. AR gene rearrangements linked to AR-V expression in CRPC.

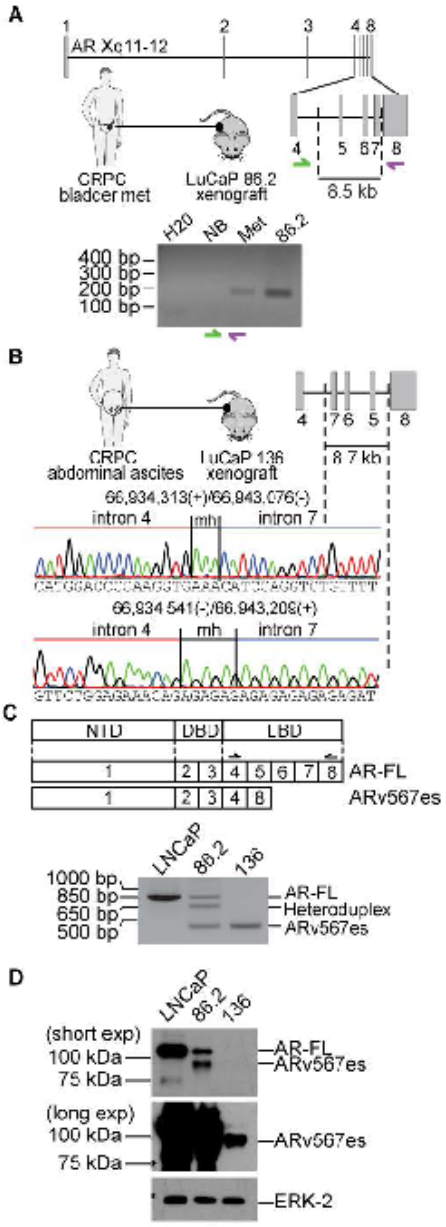


Figure 2-2. (*A*) PCR analysis of an 8.5kb deletion of AR exons 5-7 in genomic DNA isolated from the LuCaP 86.2 xenograft model, CRPC bladder metastasis used to establish LuCaP 86.2 (Met), or adjacent normal bladder (NB). H_20, water as no template control. (*B*) An 8.7 kb inversion of AR exons 5-7 in passage 2 of the LuCaP 136 xenograft, which was established from CRPC cells in abdominal ascites fluid. mh, microhomology. (*C*) RT-PCR analysis of AR mRNA in LNCaP cells, LuCaP 86.2 tissue, or LuCaP 136 tissue. Exon organization and relationship with functional protein domains for full length AR (AR-FL) and the ARv567es splice variant is shown. Heteroduplex formation in LuCaP 86.2 PCR products was confirmed in Fig. 2-5. NTD, NH_2-terminal domain; DBD, DNA binding domain; LBD, ligand binding domain. (*D*) Western blot for the AR NTD or ERK-2 (loading control) in LNCaP cells, LuCaP 86.2 tissue, and LuCaP 136 tissue. Contributions: (A,B) Dehm SM, Li Y, Hwang TH (C,D) Li Y.

Figure 2-3. LuCaP 86.2 xenograft tissue and CRPC bladder metastasis used to establish LuCaP 86.2 xenograft are heterogeneous for cells harboring the LuCaP 86.2 AR intragenic deletion signature**.**

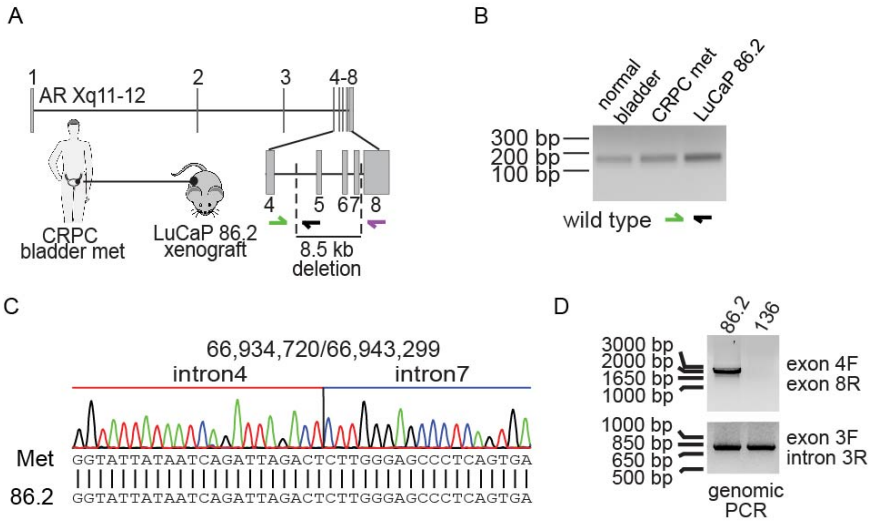


Figure 2-3. (*A*) Schematic of PCR primer design for detection of wild-type or exon 5-7 deletion AR gene architectures. (*B*) PCR testing for the presence of cells harboring normal AR gene architecture in genomic DNA from CRPC bladder metastasis, adjacent normal bladder, or LuCaP 86.2 tumor tissue. (*C*) Sanger sequence trace of the AR intragenic deletion signature in a CRPC bladder metastasis used to establish LuCaP 86.2, and alignment match with the LuCaP 86.2 deletion signature. (*D*) PCR testing for deletion of AR exons 5-7 in genomic DNA from LuCaP 86.2 and LuCaP 136 tumor tissue. For this experiment, a forward primer binding within AR exon 4 and a reverse primer binding within AR exon 8 were used. Contributions: (A- D) Li Y.

Figure 2-4. LuCaP 136 xenograft tissue is highly enriched for cells harboring intragenic inversion of AR exons 5-7.

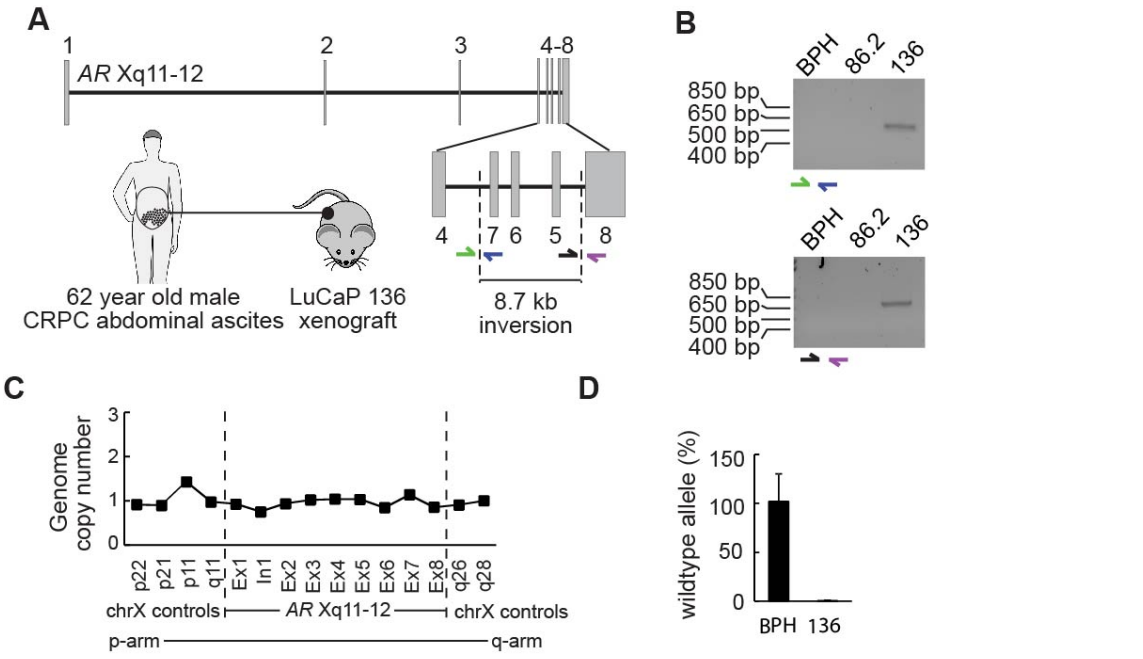


Figure 2-4. (*A*) Schematic of PCR primer design for detection of AR intragenic inversion in LuCaP 136. (*B*) PCR testing for the 5' (top) and 3' (bottom) break fusion junctions caused by intragenic inversion in genomic DNA from BPH-1 cells, LuCaP 86.2 tissue, and LuCaP 136 tissue. (*C*) MLPA genomic copy number analysis of the AR gene in genomic DNA from LuCaP 136 tissue. (*D*) Relative percentage of cells with a wild-type AR gene architecture in LuCaP 136 tumor tissue was assessed using quantitative PCR with primers specific for a wild-type, non-inverted AR gene architecture. Threshold cycle of amplification (Ct) values were calibrated relative to a control region in the AR locus (intron 1) using the relative threshold cycle of amplification method (2^{-ddCt}). Data are shown relative to BPH-1 cells, which were arbitrarily set to 100% based on the assumption that this cell line is homogeneous for cells harboring wild-type AR gene architecture in the target region. Contributions: (A- D) Li Y.

Figure 2-5. A heteroduplex PCR artifact following RT-PCR analysis of AR mRNA expression in the heterogeneous LuCaP 86.2 xenograft.

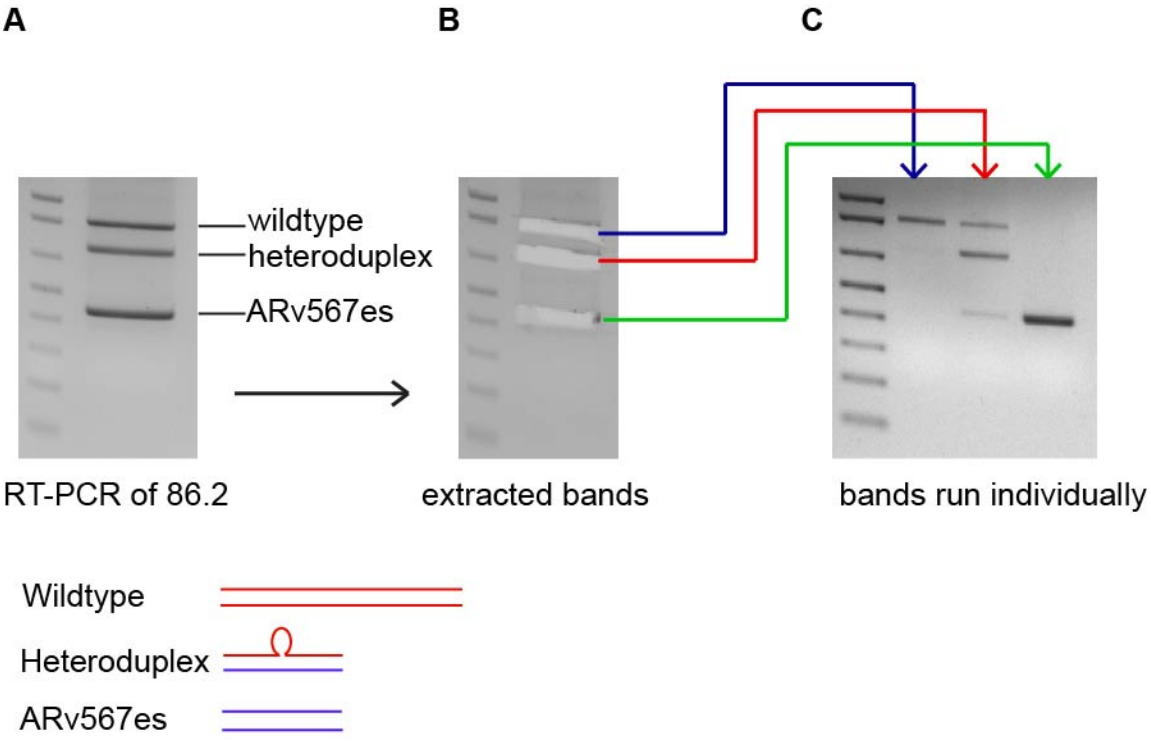


Figure 2-5. (*A*) Representative agarose gel electrophoresis of LuCaP 86.2 RT-PCR products derived using exon 4 forward/exon 8 reverse primer pairs. (*B*) Individual bands were excised, subjected to column purification, and (*C*) analyzed by agarose gel electrophoresis in separate lanes. The purified heteroduplex band was able to re-constitute all three products observed on the original agarose gel. Contributions: (A- C) Li Y.

Figure 2-6. Loss of inversion-positive cells from the LuCaP 136 xenograft following serial passage under non-castrate conditions.

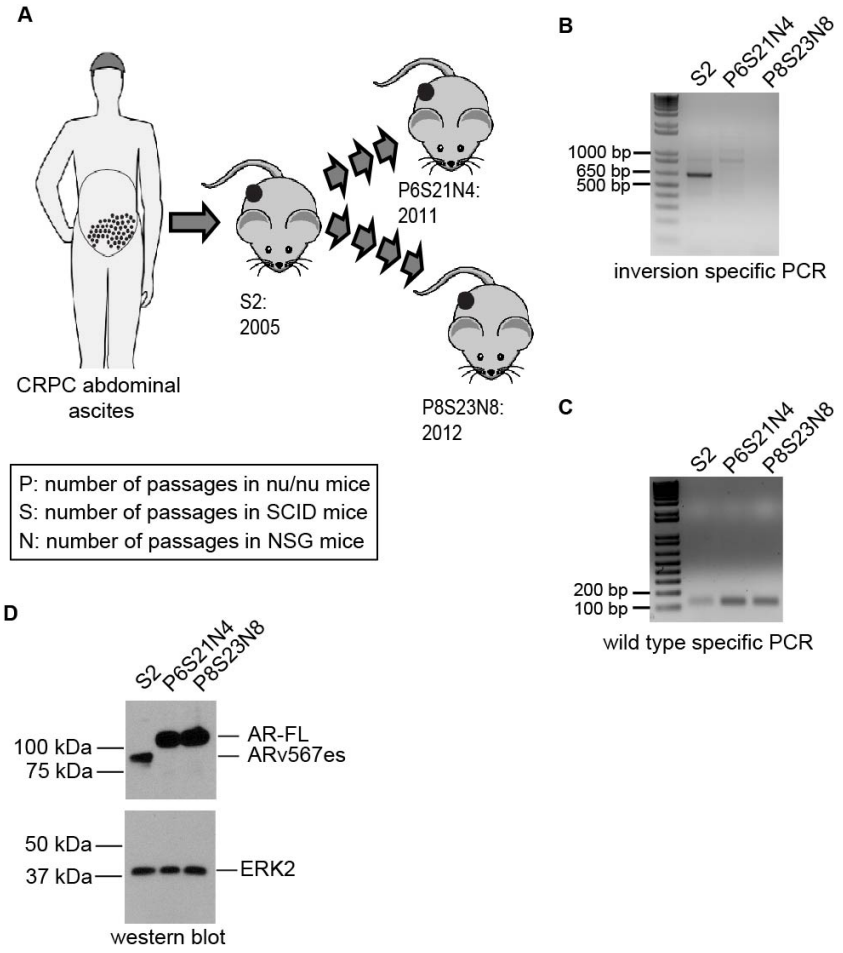


Fig. 2-6. (*A*) Origin of different passages of LuCaP 136 tumor tissue used for analysis of AR gene structure and expression patterns. (*B*) PCR testing for the 5' break fusion junction caused by intragenic inversion of AR exons 5-7 in genomic DNA from three different passages of LuCaP 136. (*C*) PCR testing for the presence of cells harboring normal AR gene architecture in genomic DNA from three different passages of LuCaP 136. (*D*) Western blot for the AR NTD or ERK-2 (loading control) in whole cell extracts from three different passages of LuCaP 136. Contributions: (A- D) Li Y.

Figure 2-7. Engineered inversion or deletion of AR exons 5-7 using TALENs recapitulates tissue-associated AR splicing events**.**

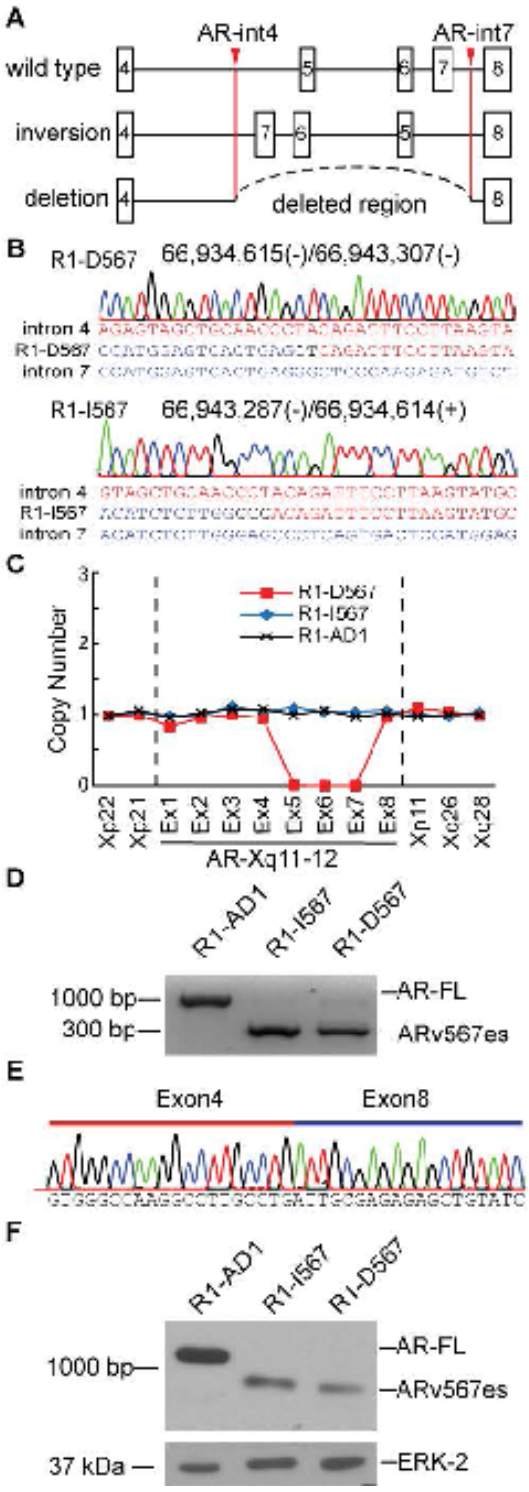


Figure 2-7. (*A*) Genome engineering strategy for creating isogenic cell lines harboring inversion or deletion of AR exons 5-7. TALEN pairs targeted to AR intron 4 (AR-int4) or AR intron 7 (AR-int7) are depicted as red arrowheads. (*B*) Representative Sanger sequencing trace of the deletion junction from the genome-engineered R1-D567 cell line, and the 5' inversion junction from the genome-engineered R1-I567 cell line. (*C*) MLPA genomic copy number analysis in parental R1-AD1 cells, and genome-engineered R1-I567 and R1-D567 cells. (*D*) RT-PCR analysis of AR mRNA in parental R1-AD1 cells, and genome-engineered R1-I567 and R1-D567 cells. (*E*) Representative Sanger sequencing trace of the AR exon 4/8 splice junction in ARv567es RT-PCR products from panel D. (*F*) Western blot for the AR NTD or ERK-2 (loading control) in parental R1-AD1 cells, and genome-engineered R1-I567 and R1-D567 cells. Contributions: (C) Hirsh B, Oseth LA.

Figure 2-8. Evaluating AR-targeted TALENs for site-specific dsDNA cleavage activity.

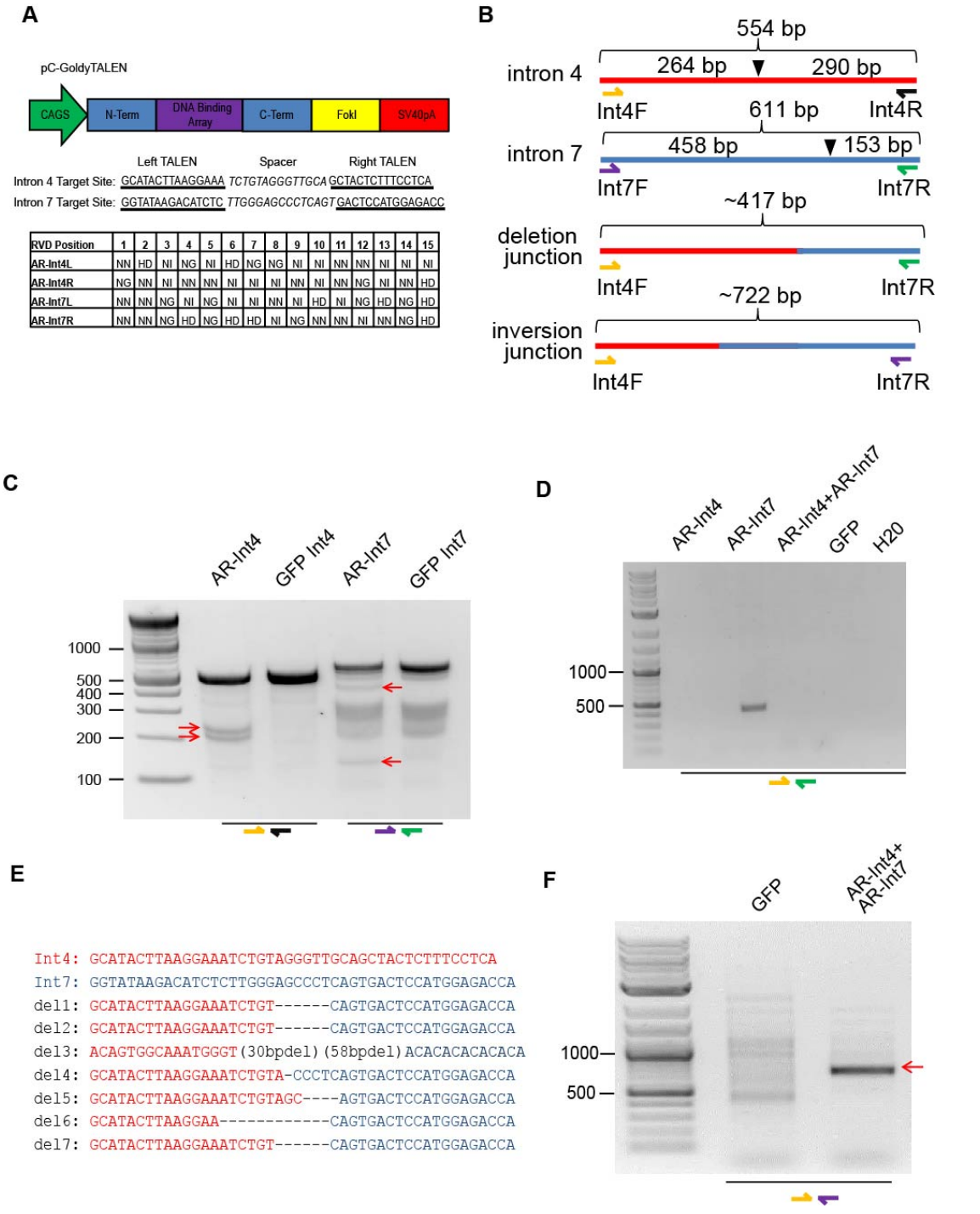


Figure 2-8. (*A*) Schematic of TALEN expression vectors (top). TALEN binding sites in AR introns 4 and 7 (underlined). Sites of targeted dsDNA cleavage are within the spacer region (italicized) (middle). Repeat-variable di-residue (RVD) module sequence of TALEN DNA binding domains (bottom). (*B*) Schematic of genomic regions surrounding TALEN target sites (black arrows). Locations of PCR primer binding sites and predicted PCR fragment lengths are indicated. (*C*) Cel1 assay, measuring indels (insertion or deletions) caused by dsDNA break repair through error prone non-homologous end joining (NHEJ), of genomic DNA from LNCaP cells transfected with individual TALENs or a GFP control vector. (*D*) PCR testing for deletion events in genomic DNA from LNCaP cells co-transfected with AR-Int4 and AR-Int7 TALENs. (*E*) Deletion junction sequences of 7 randomly picked cloned PCR products derived from PCR reactions in (*D*). (*F*) PCR testing for inversion events in genomic DNA from LNCaP cells transfected as in (*D*).

Figure 2-9. cDNA sequences of ARv567es RT-PCR products from R1-I567 and R1-D567 cells.

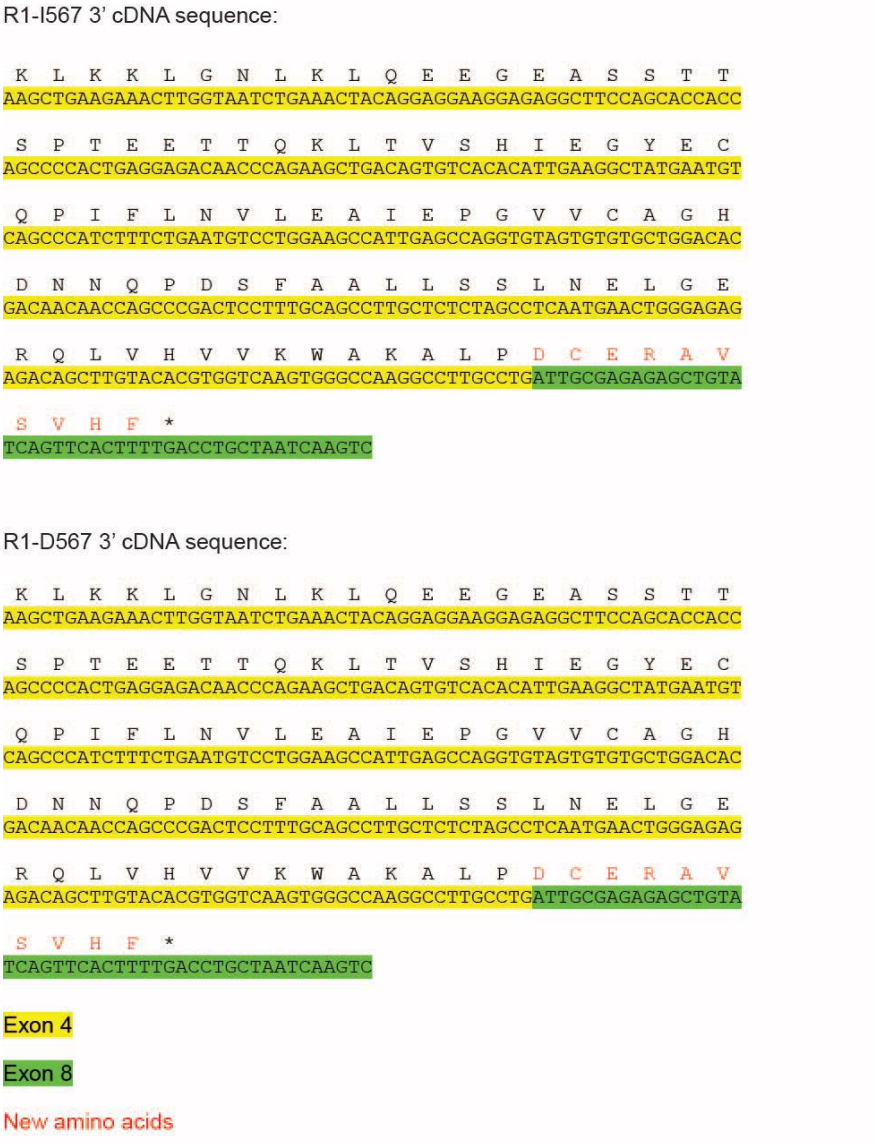


Figure 2-9. cDNA sequence derived from exon 4 is highlighted in yellow, and cDNA sequence derived from exon 8 is highlighted in green. Novel amino acids derived from the frameshift caused by AR exon 8 splicing to AR exon 4 are indicated by red letters. Asterisks denote ARv567es stop codon.

Figure 2-10. ARv567es expression induced by AR gene rearrangements drives an androgen-independent prostate cancer phenotype.

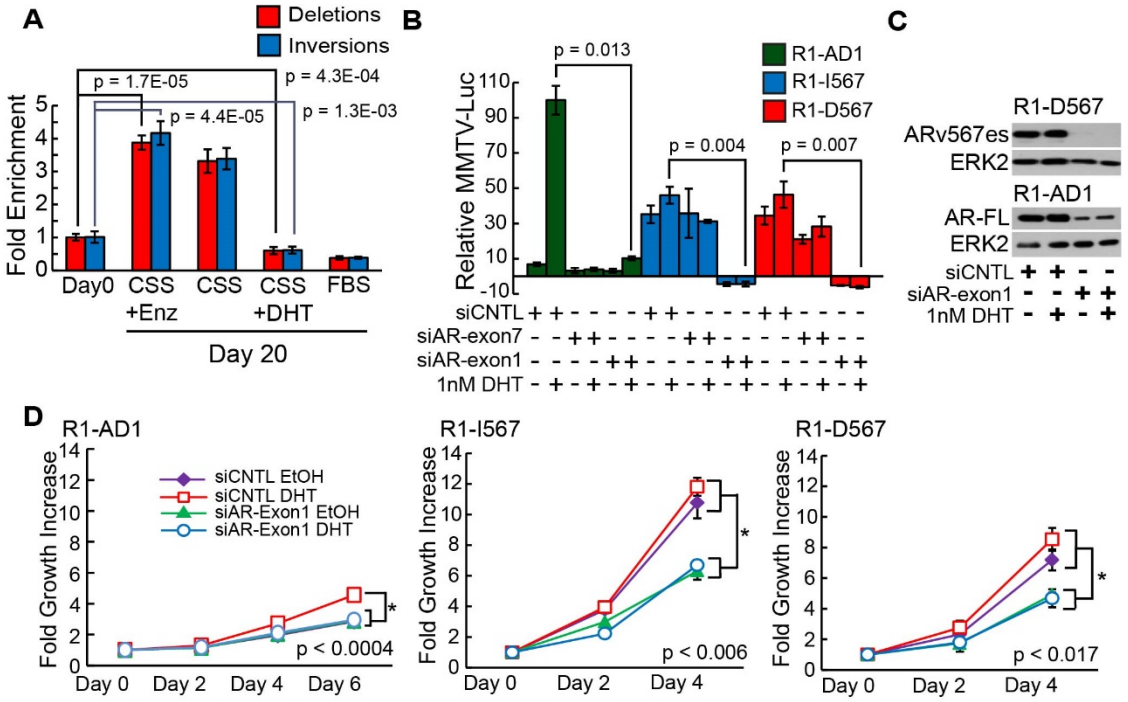


Figure 2-10. (*A*) Relative enrichment of cells with AR intragenic deletion or inversion events induced by transfecting R1-AD1 cells with AR-int4 and AR-int7 TALENs. Cells were growth to confluence after transfection, and then re-plated on Day 0 under androgen-rich (FBS or CSS + 1nM DHT) or castrate (CSS or CSS + 1uM enzalutamide/Enz) conditions. On Day 20, plates were tested for enrichment of cells with AR deletion or inversion alleles relative to all AR alleles (intron 2) by quantitative PCR (N = 6). p-values were calculated using the two-tailed *t*-test. CSS, charcoal-stripped serum; FBS, fetal bovine serum; DHT, dihydrotestosterone. (error bars = SD). (*B*) Promoter-reporter assays following co-transfection with an AR-responsive MMTV-

luciferase reporter and small interference RNAs targeted to AR exon 7 or AR exon 1. Transfected cells were treated 24h with 1nM DHT as indicated. Results are the mean (N = 3). p-values were calculated using the two-tailed *t*-test. The data are shown for one triplicate experiment, which is representative of three independent biological replicates, each of which reached statistical significance ($p < 0.05$) (error bars = SD). (*C*) Western blots for the AR NTD or ERK-2 (loading control) in R1-D567 and parental R1-AD1 cells transfected with siRNA targeting AR exon 1. (*D*) Growth assays of parental R1-AD1 cells and genome-engineered R1-I567 and R1-D567 cells transfected as in panel C (N = 4). p-values shown are the largest of any comparison between bracketed groups and were calculated using the two-tailed *t*-test. The data are shown for one quadruplicate experiment, which is representative of two independent biological replicates, each of which reached statistical significance ($p<0.05$). (error bars = SD)

Figure 2-11. ARv567es induced by AR gene rearrangements drives constitutive, androgen-independent expression of the AR transcriptional program.

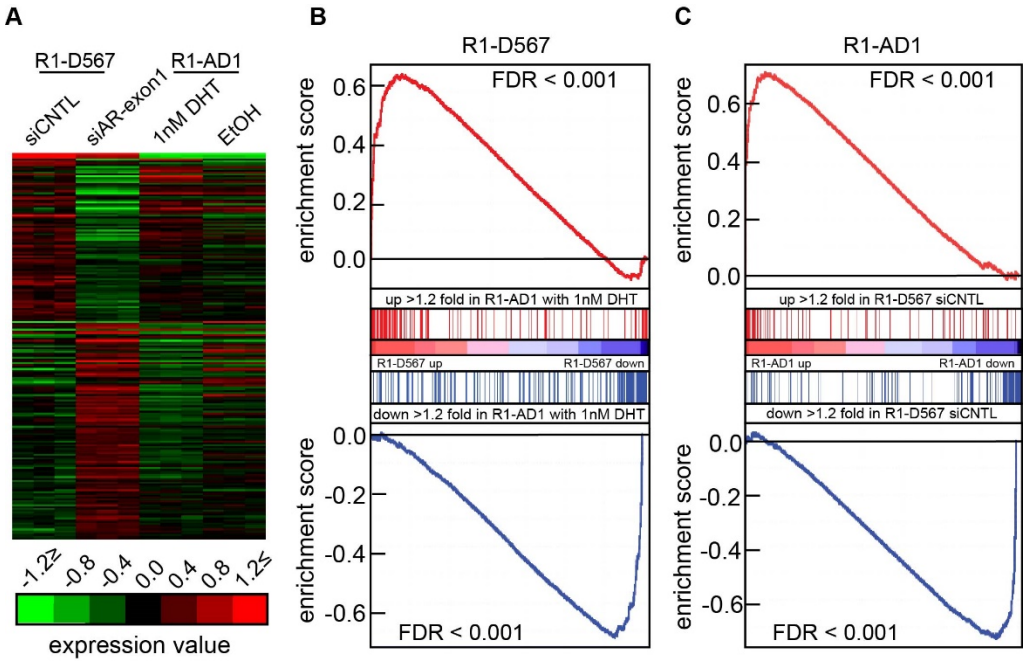


Figure 2-11. (*A*) Heat map of microarray data showing genes differentially expressed in R1-D567 cells transfected with control siRNA (siCNTL) vs. siAR targeting AR exon1 (ARv567es transcriptome), and responses of these same genes to 1nM DHT in parental R1-AD1 cells. Data represent mean centered expression changes in $\log_2$ scale for 3 independent biological replicates. (*B*) GSEA interrogating R1-D567 gene expression data for enrichment of genes that were either up regulated 1.2 fold (top in red) or down regulated 1.2 fold (bottom in blue) in R1-AD1 cells in response to treatment with 1nM DHT. False discovery rate (FDR) q-values are shown for each plot. (*C*) GSEA as in panel B interrogating R1-AD1 gene expression data for enrichment of genes that were either up regulated 1.2 fold (top in red) or down regulated 1.2 fold (bottom in blue) in R1-D567 cells transfected with control siRNA (siCNTL) vs. siAR targeting AR exon1.

Figure 2-12. Quantitative RT-PCR analysis of representative genes displaying coordinate regulation by androgen/AR and ARv567es.

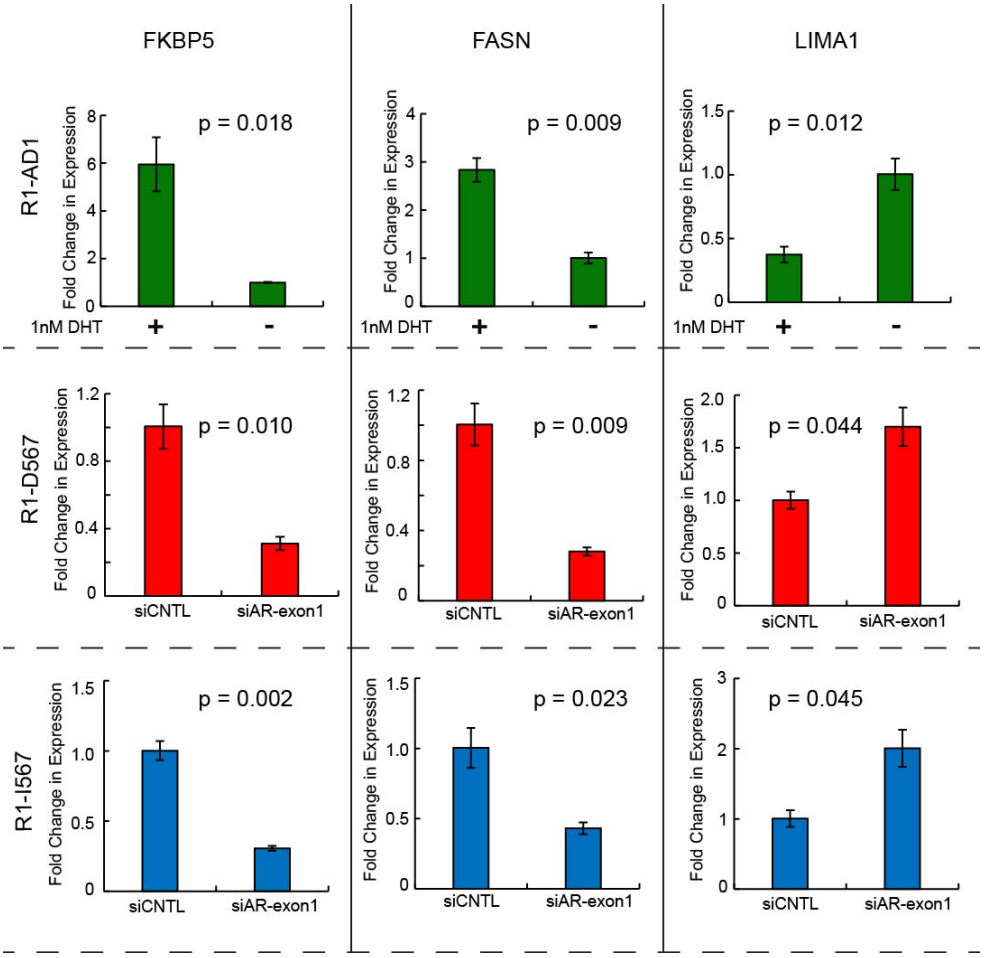


Figure 2-12. Expression of the indicated genes was assessed following androgen-treatment of R1-AD1 cells, or siRNA-mediated knock-down of ARv567es in R1-D567 and R1-I567es cells. Results are the mean (N=3). p-values shown were calculated using the two-tailed *t*-test (Error bars = SD). The data are shown for one triplicate experiment, which are representative of two independent biological replicates, each of which reached statistical significance ($p<0.05$).

Figure 2-13. Ingenuity Pathways Analysis of differentially expressed genes from microarray analyses.

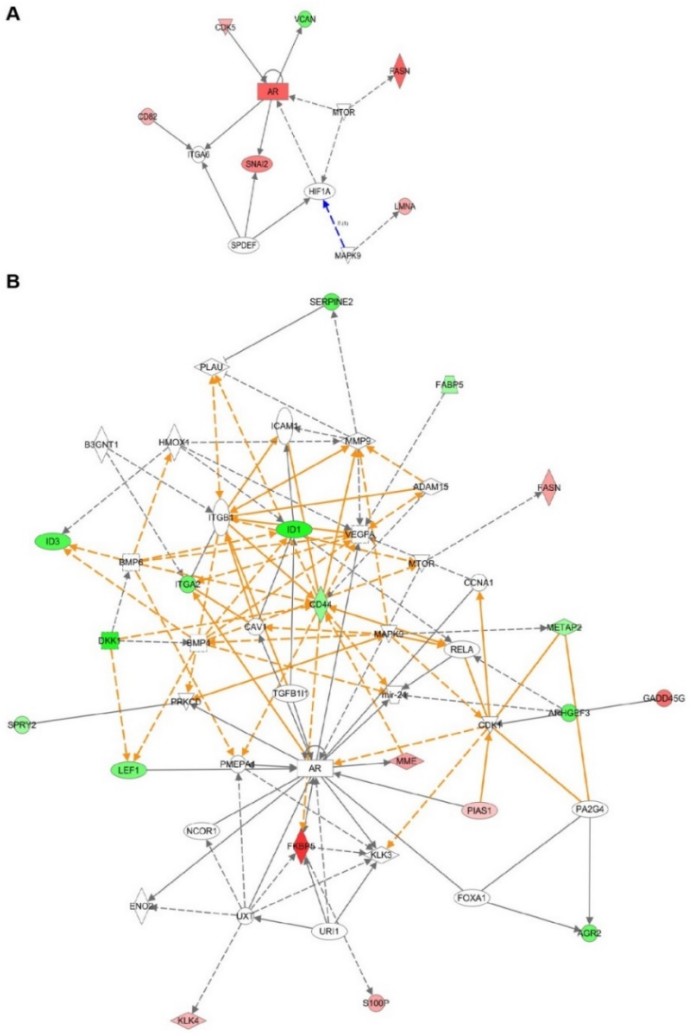


Figure 2-13. (*A*) Ingenuity Pathways Analysis was performed using genes displaying differential regulation between the ARv567es "on" state (control siRNA) and ARv567es "off" state (siRNA targeted to AR exon 1) in R1-D567 cells. (*B*) Ingenuity Pathways Analysis was performed using genes displaying differential regulation between the AR "on" state (1nM DHT) and AR "off" state (ethanol vehicle control) state in R1-AD1 cells. Contributions: Hwang TH.

2.6 Tables

Table 2-1. Primers used in chapter 2

q-PCR primers for R1-AD1 genomic DNA

qpcr-INT2 F	5'-TCCCATTTTCAGGCCTCTTA
qpcr-INT2 R	5'-GCTTAAGCCCTGGGTGGT
qpcr-INT4 F	5'-TCTGTTTCTCCAGAACAGCCTA
qpcr-INT7 R	5'-GAAATGGTCTCCATGGAGTC
qpcr-INT7 F	5'-GAGAGAGAGAGAGATGGAGTGC

LuCaP 86.2 deletion PCR

qdel F2	5’-TAGGGTTGCAGCTACTCTTTCC
qdel R3	5’-TGCTTAGCACTCAAACCCAGTA
nor F	**same as qdel F2
nor R2	5’-TGTGTGTTAGAGAGAGACAGCGA

LuCaP 136 Inversion PCR

136inv int4+	5'-ATTTGGAGTGGGTGAGTAGACTGG
136inv int7+	5'-TGCTTTTATCAGGGAGAACAGCC
136inv int4-	5'-AGCTCTCTGACTCAGACTTC
136inv int7-	5'-GTGAATGTGAAGGCACATGG

LuCaP 136 quantitative Inversion PCR

qLuCaP 136 F1	5'-CCACTTGCCTTGCCTAGAAG
qLuCaP 136 R1	5'-GGGTGGAGGAGTTGAGAACA
qAR normal R1	5'-CAAAGAAAGGCCAGTTTGGA
qAR intron1 F	5’-TGGATGGATAGCTACTCCGG
qAR intron1 R	5’-TTTACCCTGCTGAGCTGTCC

AR RT-PCR

EXON4 F	5'-GCAGCAAAGATTTCCAAACTGG
EXON8 R	5'-TGGGTGTGGAAATAGATGGG

R1-AD1 screening and cel1 PCRs

Int4 F	5'-GCAGCAAAGATTTCCAAACTGG
Int4 R	5'-CAAGCAAATTGTCCATACTGATGC
Int7 F	5'-GCTTTATCAGGGAGAACAGCCTG
Int7 R	5'-CCTCTGATTTTTGGTCTTTCAGCC

Sequencing PCRs

T7	5'-GTAATACGACTCACTATAGGG

qRT-PCRs

qpcr-GAPDH F	5'-GAAGGTGAAGGTCGGAGTC
qpcr-GAPDH R	5'-GAAGATGGTGATGGGATTTC
qpcr-FKBP5 F	5'-AGGAGGGAAGAGTCCCAGTG
qpcr-FKBP5 R	5'-TGGGAAGCTACTGGTTTTGC
FASN F1	5'-CTGAAGCGTGGCCTGAAG
FASN F2	5'-CTTCCTCACCTCCACTGAGC
LIMA1 R1	5'-TTTTGCTTGCCCATAGATCC
APIP F1	5'-ACTGGGACTGGAGGAGGAAT
APIP R1	5'-ATCACTGCACCTGCTCCTCT

Table 2-2. GSEA statistics

Name of Gene Set	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	
R1-AD1_DHT_RESPONSE_1.2_FOLD_UP	0.975728	3.733761	0	0	0	964	
R1-D567_ARV_1.2_FOLD_UP	0.732285	2.648226	0	0	0	2475	

Enriched in R1-AD1 with EtOH

Name of Gene Set	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	
R1-AD1_DHT_RESPONSE_1.2_FOLD_DN	-0.97838	-3.96718	0	0	0	649	
R1-D567_ARV_1.2_FOLD_DN	-0.74185	-2.86152	0	0	0	2539	

Enriched in R1-D567 siCNTL

Name of Gene Set	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	
R1-D567_ARV_1.2_FOLD_UP	0.961824	3.444299	0	0	0	1410	
R1-AD1_DHT_RESPONSE_1.2_FOLD_UP	0.66122	2.495106	0	0	0	3822	

Enriched in R1-D567 siAR2

Name of Gene Set	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	
R1-D567_ARV_1.2_FOLD_DN	-0.96871	-3.73744	0	0	0	1065	
R1-AD1_DHT_RESPONSE_1.2_FOLD_DN	-0.72273	-2.98245	0	0	0	3473	

Table 2-3. Gene lists used in GSEA

R1-D567_ARV_1.2_fold_UP	R1-D567_ARV_1.2_fold_DN	R1-AD1_DHT_Response_1.2_fold_UP	R1-AD1_DHT_Response_1.2_fold_DN
DAGLB	ANXA1	C13ORF15	DKK1
ERRFI1	IGSF9	ATP1A1	ETV5
ARHGAP23	PARP10	MME	KIAA1199
ATP1A1	PLEKHB1	TOMM34	FER1L3
MBOAT2	C1ORF116	MBOAT2	RAB7B
CYBA	MB	UGT2B7	ANXA1
CYP27B1	CYFIP2	NT5DC3	AGR2
GNPDA1	LPAR5	PRR15L	ID1
PLEKHA6	HS.370359	HOMER2	IGSF9
CEBPD	ZMIZ2	HERC5	MYOF
HS.538962	ABCA1	CHN2	FOXC1
BTG2	N4BP2L2	UGT2B11	KIAA1671
FASN	ALDH3A2	PYGB	ETV4
TSPYL2	SIPA1L2	ADAMTS9	KIAA0040
GSTO1	LOC100132901	SOCS2	FGFBP1
ACOT7	ID2	UGT2B28	FAM108C1
TSPAN18	HOXC13	PLEKHA6	SPRY2
OSBP	ZNF323	CEBPD	ENC1
RDH11	CYP3A5	HS.370359	OCIAD2
LOC100133185	TSPAN8	FASN	SPRED2
TINP1	FAM134B	TSPYL2	HS.434957
ACSS2	FLJ22184	TIPARP	MAP7
LOC653375	LRP10	S100P	C20ORF118
BTBD7	TP53INP1	LOC93622	LPAR5
HSPB3	VCAN	LOC729768	FAIM2
GLB1	C9ORF152	MVD	INO80C
AR	C1ORF115	ERN1	VWF
TTLL12	EPCAM	MFSD2	DUSP6
PIGP	CLDN3	C16ORF79	PHCA
APOD	SSPO	PACSIN1	TM4SF1
FLJ25404	GSN	CMAS	AKR1B10
LMNA	OXTR	APIP	PHLDA1
LRRC8A	SCNN1A	ETNK2	CMBL
DAPL1	XBP1	TAP1	EMP1

TABLE 2-3 cont.			
R1-D567_ARV_1.2_fold_UP	R1-D567_ARV_1.2_fold_DN	R1-AD1_DHT_Response_1.2_fold_UP	R1-AD1_DHT_Response_1.2_fold_DN
KDELR2	LOC653108	LOC284422	LRRC49
RAG1AP1	DACH1	TTLL12	CD58
TUBA3D	FYCO1	PDIA5	RBM47
EHD4	LOC645381	EPHX2	ALDH3A2
SAFB2	AARS	TMEM38B	CPNE4
LIPH	HS.155736	SHRM	CEACAM1
DUOX1	CXADR	LRRC8A	AGPAT9
NSA2	SERPINI1	EXOSC5	CLIC1
PMM1	ASS1	C19ORF48	PRSS23
BOLA3	ADNP	TNS3	ARHGEF3
SRP68	NRCAM	KDELR2	SIPA1L2
LOC100128196	ALDH1L2	TUBA3D	ID2
HSPB6	CBLN2	SGK1	ALDH1A3
HS.564874	ABR	SAFB2	TGFB3
ODF2	BRI3	COPS8	HOXC13
TSC2	FAM84B	PECI	CAMP
FBXO18	HS.28456	FAM174B	VAV3
CD82	ANKRD50	BIN1	SOX7
C5ORF25	PCMTD1	PMM1	STK3
LGI2	CEBPG	SRP68	LOC100134170
IQCK	LOC728533	CUX2	TSPAN8
ZBTB16	GAS6	TSPAN33	C1ORF127
DGUOK	STX3	BBS10	MFSD6
PALLD	CDH18	NUBP2	ADAMTS1
AZI1	ZNF615	TSC2	PCDHB2
IPO7	ATP9A	TBC1D4	CCDC3
CDK5	DMAP1	SLC4A7	BCL3
ISCA1	HIST1H3H	FLJ90086	PHLDA2
SSSCA1	ASAP2	RABAC1	C9ORF152
ABHD6	MBP	AFF3	TMTC1
CDC42	GSTA4	KLK4	C1ORF115
CD99L2	AVIL	RALY	CREB3L4
RASD1	ZNF524	FADS1	FUT4
LOC100132658	HLX	FKBP5	AMHR2
TKTL1	RPS6KA2	ANXA2	CTSL2
CCDC6	WASL	LDHB	GSN

TABLE 2-3 cont.			
R1-D567_ARV_1.2_fold_UP	R1-D567_ARV_1.2_fold_DN	R1-AD1_DHT_Response_1.2_fold_UP	R1-AD1_DHT_Response_1.2_fold_DN
ELMOD1	HS.492187	SSR3	SHROOM2
AP1S2	SDC4	HS.571245	CLCF1
PDLIM7	FAM102A	RNF150	PTPN13
ABCG4	GSTA3	LIG1	OXTR
SS18	ZDHHC1	FZD9	NDE1
JPH1	ZNF467	BEST2	LOC653108
FAM96B	CD47	TMEM214	NCAM2
SH2B3	ZFAND1	PDLIM1	PLEKHG4
NLRX1	GPNMB	CDK5RAP3	RIN2
DUSP3	LOC653506	GADD45G	TNFRSF19
CAMKV	PCDH19	EAF2	TXNDC5
FABP6	CPNE3	CYB5R3	ZBTB33
SNAI2	TGFBR3	SLC6A3	HS.25318
SIX4	HS.559604	ABHD6	DNAJC12
SNORA67	PIK4CA	WDR41	HS.155736
UNC5A	MACROD1	GAA	CXADR
NOC2L	RALYL	NUCB1	CARM1
LAPTM4B	CDC42SE2	ABCC4	ZMYND15
GAGE12J	ZBTB42	CECR6	SLC12A2
DKK3	AFF2	HS.19339	ITGA2
	DNAJB9	LONRF1	CRABP2
	P4HTM	BTBD11	SLC20A2
	ARHGAP24	MRFAP1	DNAJC22
	KAT2B	C6ORF81	ACOX2
	AKTIP	FHDC1	ASS1
	ZSCAN16	GMPPB	LAD1
	RWDD2A	ARSA	CXORF26
	SIGIRR	RHOB	LIMA1
	MAL2	GMPPA	MPZL2
	HS.335413	LACTB	CYTH2
	SLC1A5	KLHDC4	ABR
	RNF149	CADPS2	FAM84B
	SLC23A3	PHACTR3	LYPD6B
	HS.352818	NTHL1	TMSB4X
	SARS	CYB5A	SCGB2A2
	ADM2	EEF2K	SYTL2

TABLE 2-3 cont.			
R1-D567_ARV_1.2_fold_UP	R1-D567_ARV_1.2_fold_DN	R1-AD1_DHT_Response_1.2_fold_UP	R1-AD1_DHT_Response_1.2_fold_DN
	C11ORF54	GHR	PROM2
	FOXN2	STK39	CCDC50
	HS.444683	SOD2	LEF1
	NAGK	LOC387856	TNFRSF21
	ILDR1	TXNDC16	ID3
	GOLSYN	SAMM50	ANKRD57
	TMEM125	IMPA2	CAPRIN2
	KLHDC8B	VPS26B	IPO8
	RP1L1	FBXL15	TMBIM4
	PPP1R9A	MRPL2	DLX3
	WDR20	TRAP1	JMJD1C
		NEURL1B	SCGB1D2
		FLJ10324	LMCD1
		SNX25	ACYP2
		HTRA2	GSTA4
		PLOD1	GPR177
		PACS1	NFKBIZ
		PIAS1	COBLL1
		RNF103	VIPR1
		SLC25A1	NAB1
			SDC4
			C6ORF117
			C9ORF3
			ANKRD22
			GJB4
			FRMD6
			GSTA3
			C15ORF52
			AHR
			LOC730413
			CYP4X1
			TFAP2A
			NRP1
			TLE1
			PNPLA7
			ZNF428

TABLE 2-3 cont.
R1-AD1_DHT_Response_1.2_fold_DN
ZCCHC12
SERPINE2
KIAA1324
CCDC102A
BBS2
CD44
FAM149B1
PLEKHA9
SLC44A2
C12ORF26
PLEKHH2
MARCKS
GPX2
FES
CDH3
SNCG
IGSF3
SELENBP1
FABP5
MIDN
CNKSR3
METAP2
WDR72
SOX2
EFHD1
C1QL4
HOXB5
ATP2C1
DIP2B
IRX5
C10ORF32
CASK
GNRHR
LPHN2
GPR56
RBM22

TABLE 2-3 cont.
R1-AD1_DHT_Response_1.2_fold_DN
PTP4A2
RNY5
LOC642956
LITAF
LOC100128098
IFNGR2
SYNGR4
SETMAR
FLJ40125
B9D2

CHAPTER 3: TALEN-mediated modeling of AR gene rearrangements associated with AR-V7 production in human prostate cancer cell lines.

Prostate cancer is the most frequently diagnosed male cancer. Therapies that target the androgen receptor (AR) are highly effective for management of aggressive disease. However, development of therapeutic resistance and emergence of the lethal castration-resistant prostate cancer phenotype is frequent. One mechanism of therapeutic resistance is the acquired expression of constitutively active AR splice variants (AR-Vs). However, the mechanism by which they are expressed remains poorly understood. Overexpression of splice variant AR-V7 has been associated with intragenic *AR* rearrangements but it is not known whether specific rearrangements are causative of overexpression. In this study, we induced sequence-specific nuclease-mediated gene modifications to model intragenic *AR* rearrangements that are associated with AR-V7 over-expression in castration-resistant prostate cancer (CRPC) cell lines 22Rv1, and CWR-R1. Our data demonstrate that duplication of a ~35kb region surrounding exon 3, as in 22Rv1, can enhance AR-V7 expression. However, low levels of AR-V7 persisted even after the duplication was reversed. This suggests that there are multiple, non-mutually exclusive modes of AR-V7 overexpression.

3.1 Introduction

Prostate cancer (PCa) is the most frequently diagnosed cancer in men and the second highest cause of male cancer deaths (Siegel et al, 2013). Prostate tissue homeostasis is maintained by the transcriptional activity of the AR. And AR continues to function as a master regulator in prostate cancer (Garraway and Sellers, 2006). AR

consists of an amino-terminal transcriptional domain (NTD), a DNA binding domain (DBD), and a carboxyl-terminal ligand binding domain (LBD). When not bound to ligand, AR remains in an inactive state in the cytoplasm. Upon binding to dihydrotestosterone (DHT), AR translocates to the nucleus where it dimerizes and activates transcription (Callewaert et al, 2006; Christiaens et al, 2002; He et al, 2004; Jenster et al, 1995). Accordingly, when PCa progresses to an aggressive, metastatic disease, the primary treatment is androgen deprivation therapy (ADT), which is achieved by reducing the availability of androgens (castration) or use of chemical antagonists that bind to the AR LBD (Ryan and Tindall 2011). Inevitably, PCa develops resistance to these therapies and progresses to castration resistant prostate cancer (CRPC) within several years.

ADT resistance is often caused by reactivation of AR signaling though myriad alterations in AR regulation. The best defined AR reactivation mechanisms include point mutations or amplification of the *AR* gene, or intratumoral androgen biosynthesis (Shafi et al, 2013; Chang et al, 2013). These alterations have been re-retargeted successfully in the clinic with the anti-androgen enzalutamide as well as the androgen biosynthesis inhibitor abiraterone acetate. However, the overall survival benefit achieved by these AR re-targeting strategies is limited to a few months (Scher et al, 2012; de Bono et al, 2011). An additional AR reactivation mechanism that occurs in CRPC is expression of AR splice variants (AR-Vs) (Dehm and Tindall, 2011). AR-Vs, such as AR-V7 and AR-V567es, contain the NTD and DBD of the AR protein, but are prematurely truncated prior to the ligand binding domain, obviating the requirement for DHT. In this way,

variants effect a constitutive AR signaling program and are inert to LBD targeted therapies (Li Y et al, 2013; Nyquist and Dehm, 2012).

Understanding the modes of AR-V expression is important for precision medicine, prognostic biomarker research, and drug development. One study found that splice factors U2AF65 and ASF/SF2 are required for splicing of AR cryptic exon 3 (CE3), the exon in AR-V7 (Liu LL et al, 2013). However, the study did not address how these splicing factors are involved in the inappropriate expression of AR-V7. Intragenic AR rearrangements have been found in cell lines where high levels of AR-V expression drive castration resistant growth (Li Y et al, 2011; Li Y et al, 2012; Li Y et al, 2013).

One example of this scenario is the CRPC 22Rv1 cell line, which is characterized by a ~35kb intragenic tandem duplication in the *AR* gene and high-level expression of several discrete AR-V species, including AR-V7 (Li Y et al, 2011). Similarly, CWR-R1 has been shown to be a heterogeneous cell line, harboring a subset of cells that are characterized by a ~50kb deletion within *AR* intron 1 and high-level expression of AR-V7 (Li Y et al, 2012 and 2013). These findings strongly implicate AR gene rearrangement as the common mechanism for functional AR-V7 expression in these cell lines, but this has not been established experimentally. To determine whether the intragenic AR rearrangements are sufficient to drive high level AR-V7 expression, we used Transcription Activator-Like Effector Nucleases (TALENs) to model AR-V7 associated rearrangements in an isogenic context (Cermak et al, 2011).

3.2 Results

3.2.1 Genetic correction of *AR* in 22Rv1

The CRPC cell line, 22Rv1, harbors a duplication in a region of *AR* that contains exon three as well as the cryptic exons involved in the generation of AR-Vs. In order to determine the significance of this duplication to AR-V expression, we designed a TALEN pair, designated tAR2, to bind within the duplicated region (Fig. 3-1 *A*; Tables 3-2, 3-3). We reasoned that simultaneous double-strand breaks, one within and one outside the duplication, would delete the intervening sequence, thereby generating a virtually normal *AR* gene structure.

First, we confirmed the activity of tAR2 with a PCR-digest assay which measures non-homologous end joining (NHEJ) mutations introduced within the cleavage site located in the spacer sequence between the two TALEN binding sites. In this instance, an Afl III endonuclease site was located in the spacer; error prone NHEJ resulting from tAR2 induced dsDNA breaks mutated the Afl III site, causing a PCR product spanning this site to be resistant to Afl III digestion (Fig. 3-1 *A*). PCR of genomic DNA derived from cells expressing tAR2 revealed high levels of mutations derived from NHEJ compared to genomic DNA from control cells expressing a vector encoding green florescent protein (GFP) (Fig. 3-1 *B*). The findings confirm efficient, site-specific nuclease activity of tAR2.

Next, 22Rv1 cells were transfected with tAR2 and plated at limiting dilution in order to obtain clonal populations of cells with an excised duplication. Subclones were screened by PCR-digest of the tAR2 cutsite as performed in (Fig 3-1 B); those with a wild type band were excluded from further screening. Clones that did not display wild-type PCR-digest signal were tested for the presence of the duplication using a multiplex ligation-dependent probe assay (MLPA). This approach yielded four discrete clonal cell

lines (B7-5, B34-4, C47-2, and E43-2) that displayed loss of the duplicated genomic segment. One clone, B7-5, could not be characterized further due to an abnormally high rate of cell death in culture.

3.2.2 Recreating the CWR-R1 intragenic *AR* deletion in R1-AD1

The CWR-R1 cell line harbors a 50kb deletion in intron 1 that is associated with AR-V7 overexpression. In order to establish the role of this rearrangement in AR-V7 production, we designed a strategy to recreate this deletion in R1-AD1, a sub-line of CWR-R1, which contains a structurally normal *AR* gene and expresses very low levels of AR-V7 (Li Y, 2013). Two TALEN pairs, tAR1-1 and tAR1-2, were designed to created double strand breaks flanking the ~50kb region in intron 1 associated with AR-V7 overexpression (Fig. 3-2 *A*). Coordinate action of the TALEN pairs induced two double strand breaks which were repaired by NHEJ to create a deletion as detected by deletion-specific PCR (Fig. 3-2 *B*). When tAR1-1 and tAR1-2 were transfected separately, no deletions were detected (Fig. 3-2 *B*). Next, tAR1-1 and tAR1-2 expressing R1-AD1 cells were plated at limiting dilution to isolate subclones harboring the deletion. Clones were then screened for the presence of the deletion as well as cellular purity via genomic PCR using primers pairs designed to discriminate between cells with an unmodified AR gene locus and cells harboring the targeted deletion.

Three clones were identified that were positive for the AR intron 1 deletion and negative for a wild-type (WT) AR PCR signal. These discrete cell lines were designated R1-dint1a-c (deletion of intron 1) (Fig. 3-2 *C*, *D*). Clones R1-dint1b and R1-dint1c were

confirmed to harbor a deletion in intron 1 by MLPA (Fig.3- 2 *E*). However, R1-dint1a was positive for the MLPA probe signals corresponding to the deleted region. We interpreted these data to indicate that the excised AR intron 1 segment had likely integrated elsewhere in the genome. Consequently, R1-dint1a was excluded from further analysis.

3.2.3 Effects of intragenic *AR* rearrangements on AR-V expression.

Next, western blots were used to determine whether the induced rearrangements affected AR-FL and AR-V protein levels. As AR-FL and AR-V levels may be influenced by androgens, we assessed protein expression in steroid-depleted medium supplemented with androgens or antiandrogens. Compared with parental 22Rv1 cells, subclones engineered using tAR-2 expressed much lower levels of AR-V7 than 22Rv1, suggesting that the duplication is necessary for high levels of AR-V expression (Fig. 3-3 *A*). Additionally, full length AR (AR-FL) levels generally increased under androgenic conditions, whereas AR-V levels generally increased under androgen depleted conditions.

22Rv1 also expresses an AR-V, consisting of exons 1-3 and cryptic exon 2b, termed 1/2/3/2b (Dehm et al, 2008). We performed an immunoprecipitation with an 2b specific antibody followed by an immunoblot with an amino-terminal AR antibody in order to detect the presence of any 2b containing variants. As expected, reversing the duplication eliminated all 1/2/3/2b expression from the modified clones (Fig. 3-3 *B*).

Contrary to expectations, recreating the CWR-R1-associated deletion in R1-AD1 cells did not induce over-expression of AR-V7 (Fig. 3-3 *C*). In contrast with the high AR-V7 levels displayed by CWR-R1, R1-AD1 and the subclones all had very low expression of AR-V7. This indicates that the intron-1 deletion is not sufficient for AR-V7 overexpression.

3.2.4 Effects of intragenic *AR* rearrangements on growth

To characterize the androgen responsive growth of the modified cells, crystal violet growth assays were performed in CSS medium containing either 1nM DHT or ethanol vehicle. The 22Rv1 derived lines E43-2 and B34-4 retained androgen responsive growth (Fig. 4 *A*). In contrast, C47-2 demonstrated androgen-mediated growth repression under these culture conditions. Notably, all cell lines demonstrated some level of castration resistant growth. This was expected since there was significant expression of AR-V7 (Fig. 3 *A*). Additionally, other signaling molecules, such as epidermal growth factor receptor (EGFR) (Ateeq et al, 2011), have been shown to contribute mitogenic signals in 22Rv1.

As expected, R1-AD1 and its derived clones displayed androgen responsive growth as well as a slower rate of growth under androgen depleted conditions (Fig. 4 *B*). R1-dint1b and R1-dint1c did not recapitulate the androgen independent growth of CWR-R1, which further suggests the intron 1 deletion is not sufficient to engender an AR-V driven mechanism of castration resistance.

3.2.5 AR-V7 contributes to full-length AR signaling and castration resistant growth

Since the 22Rv1 subclones had measureable castration resistant growth, we sought to establish whether residual AR-V7 expression in the subclones contributed to castration resistant cell growth. To this end, we measured the relative output of AR transcription with a 4xARE-Luciferase reporter in cells co-transfected with siRNAs that selectively target full-length vs AR-V7 expression. Interestingly, AR-V7 contributed to the strength of DHT-induced AR-FL signaling (Fig. 5 *A,B*). When the non-targeting siRNA control group (siCNTL) was compared to the AR-V7 targeting group (siCE3), knockdown of AR-V7 decreased luciferase expression under androgenic conditions. Interestingly, knockdown of AR-V7 had no effect on luciferase expression in the 22Rv1 subclones. In contrast, AR-V knockdown in 22Rv1 reduced luciferase expression.

To establish the relative roles of AR-V7 and AR-FL in driving castration resistant growth, we performed siRNA knockdowns of either AR-FL or AR-V7 in the 22Rv1-derived cell lines. The growth of these cells in CSS medium was then measured using crystal violet assays. As expected, siRNA knockdown of AR-V7 significantly reduced the growth of 22RV1 and its subclones (Fig. 5 *C*). AR-FL also significantly contributed to growth in B34-4, E43-2 and, to a lesser extent, 22Rv1. Interestingly, when growth response to androgen stimulation was compared in cells transfected with control siRNA or siRNA targeting exon CE3, a greater dynamic response was observed when AR-V7 expression was silenced (Fig. 5 *D*). This indicates that AR-V7 may compensate for sub-optimal AR-FL signaling in a dose dependent manner and thereby enhance castration resistant growth.

3.3 Conclusions

In summary, we demonstrated that a ~50kb deletion in intron 1 of AR, as seen in CWR-R1, is not-sufficient to drive high-levels of AR-V7 expression. However, our data indicate that a duplication of a ~35kb region containing exon 3 and cryptic exon CE3 enforces high-level expression of AR-V7 and 1/2/3/2b. Intriguingly, significant basal expression of AR-V7 in the gene corrected clones indicates that another mechanism of AR-V7 over-expression is present in this model system. Even though the 22Rv1 subclones had lower levels of AR-V7 expression, AR-V7 still contributed to AR signaling and CRPC growth. Importantly, AR-V expression patterns in the gene corrected 22Rv1 clones are similar to levels seen in patient-derived CRPC tissues (Jernberg et al, 2013; Hornberg et al, 2011).

The mechanism of basal AR-V7 expression is unknown. One possibility is that AR-V7 expression is related to global changes in alternative splicing commonly found in cancers (Bonomi et al, 2013). Splice factors U2AF65 and ASF/SF2 were shown to bind to a splice enhancer site that regulates AR-V7 expression (Liu et al, 2013). However, it is unclear how these factors are involved in the pathogenic overexpression of AR-V7. Therefore, further characterization of the expression and post-translational modification of U2AF65 and ASF/SF2 in AR-V7 overexpressing cell lines is necessary for establishing their role in exon CE3 splicing.

3.4 Methods

Multiplex Ligation-dependent Probe Amplification: Assays were performed as described previously (Li Y et al, 2012).

Cell growth assays: Crystal violet growth assays were performed as previously described (Li Y et al, 2011). Briefly, cells were electroporated with 200 pmoles siRNA as previous described (Li Y et al, 2013). The electroporated cells were plated in 24 well plates at 2.5*10^4 cells per well in RPMI 1640 supplemented with 10% charcoal stripped serum (CSS). After 24 hours, media was replaced with RPMI + 10% CSS supplemented with 1nM DHT or vehicle control (ethanol).

Genomic DNA PCR: Genomic DNA was isolated using Nucleospin tissue kit (Catalog 740952, Machery-Nagel). Genomic DNA was amplified using Phusion-HF (M0530S New England Biolabs, Ipswich, MA) according to the manufacturer's protocol. Primers used in PCR reaction are listed in Table 3-1. Products were cloned and sequenced as previously described (Nyquist et al, 2013)

Genomic PCR digest:

Genomic DNA was isolate and amplified as described earlier (see Genomic DNA PCR). Genomic PCR products were column purified with Qiaquick PCR purification kit (28104 Qiagen. Valencia, CA). Next, 500ng of PCR product was then digested with 10 units of Afl III (R0541S New England Biolabs, Ipswich, MA) for 2 hours then run on an agarose gel.

Construction of TALENS: TALEN construction was carried out as previously described. (Cermak et al, 2011; Nyquist et al, 2013)

Immunoprecipitation and Western Blots: Immunoprecipitations and western blots were performed as previously described (Li Cancer Res 2011; Chan et al, 2012). Antibodies used in this study include: anti-AR NTD polyclonal (N-20, Santa Cruz Biotechnology), anti-ERK2 monoclonal (D-2), Santa Cruz Biotechnology), anti-AR NTD monoclonal (AR-441, Santa Cruz Biotechnology), anti-AR-V7 monoclonal (AG10008, Precision Antibody) and a polyclonal antibody specific for the COOH-terminal extension encoded by AR exon 2b (Chan et al, 2012).

Lucifierase Assay: Dual luciferase assays were performed with a 4X-ARE luciferase reporter (A gift from Dr. Michal Carey, University of California, Los Angeles) and a SV40-renilla reporter (Promega) for transfection control as previously described (Dehm et al, 2008)

Cell lines and Cell Culture: Androgen-responsive R1-AD1 prostate cancer cells with a structurally normal AR gene locus has been described (Li Y et al, 2013; Nyquist et al, 2013). Cell lines 22Rv1 (CRL-2505) and LNCaP (CRL-1740) were obtained from ATCC (Manassas, VA, USA). Cells were maintained in RPMI 1640 with 10% FBS in a 5% CO_2 incubator at $37^{O}C$

Generation of Cell Lines: To generate R1-AD1 derived 48kb intron 1 deletion lines, TALENs tAR1-1 and tAR1-2 were electroporated into R1-AD1 as previously described (Nyquist et al, 2013) and allowed to recover in RPMI 1640 media containing 10% FBS. Cells were single cell cloned by seeding at limiting dilution in 10cm dishes. Cells were

screened using a deletion specific PCR with primer pairs ARint1 F1 and ARint R2. To generate 22Rv1 subclones without a duplication, cells were electroporated and seeded at limiting dilution as described above. Next, cells were screened with PCR-digest to detect an intact Afl III binding site in the tAR2 spacer sequence. Clones without an intact Afl III were further screened by MLPA.

3.5 Figures

Figure 3-1 TALEN mediated genetic correction of the duplication in 22Rv1.

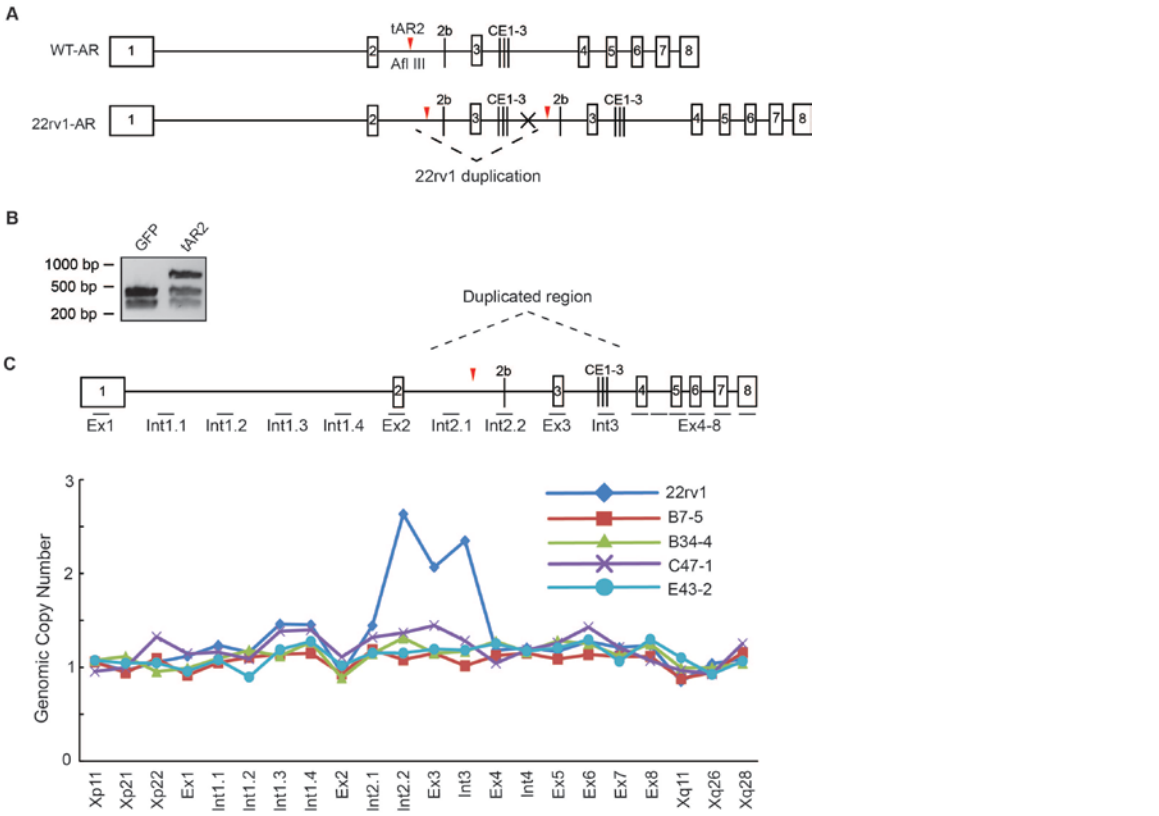


Figure 3-1 (*A*) Gene structure of wild type AR (WT-AR) in comparison to that of 22Rv1. Cryptic exons are marked 2b and CE1-3. Red arrows mark the tAR2 TALEN binding sites. The endonuclease target site of Afl III is present at the tAR2 binding site (*B*) PCR digest indicating tAR2 induced mutations. The tAR2 cutsite was amplified by PCR and digested with Afl III. TALEN activity followed by NHEJ repair induces mutations in the Afl III binding site which is indicated but an uncut PCR product of ~700bp. GFP is the negative control. (*C*) MLPA of AR locus of 22Rv1 and its subclones. AR gene locus is shown with MLPA probe sites. Primer sequences are listed in TABLE 3-1.

Figure 3-2 Modeling the CWR-R1 *AR* gene rearrangement in R1-AD1.

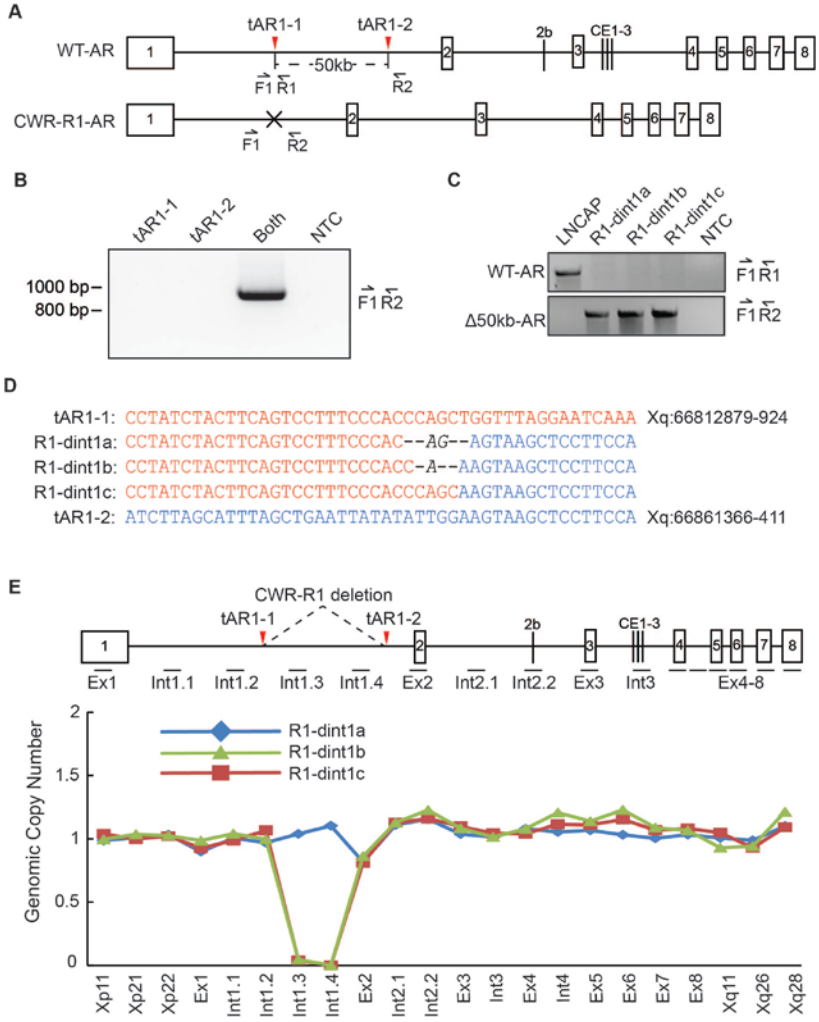


Figure 3-2 (*A*) Gene structure of wild-type AR in comparison to that of CWR-R1. The TALEN pairs tAR1-1 and tAR1-2 are indicated with red arrows. Primer binding sites F1, R1, and R2 are indicated with black arrows. (*B*) Deletion spanning genomic PCR with primers designed to detect tAR2-mediated intragenic deletion in AR intron 1. (*C*) Genomic PCR with primers designed to discriminate between clonal sublines harboring an unmodified AR locus and clonal sublines harboring tAR2-mediated intragenic deletion in AR intron 1.(*D*) Genomic sequences of the deletion break fusion junctions in clonal sublines derived from tAR2-mediated genome engineering. (*E*) MLPA-based interrogation of genomic copy number at distributed sites along the length of the AR genein clonal sublines derived from tAR2-mediated genome engineering. Primer sequences are listed in TABLE 3-1.

Figure 3-3 Rearrangement driven effects on AR-V7 expression

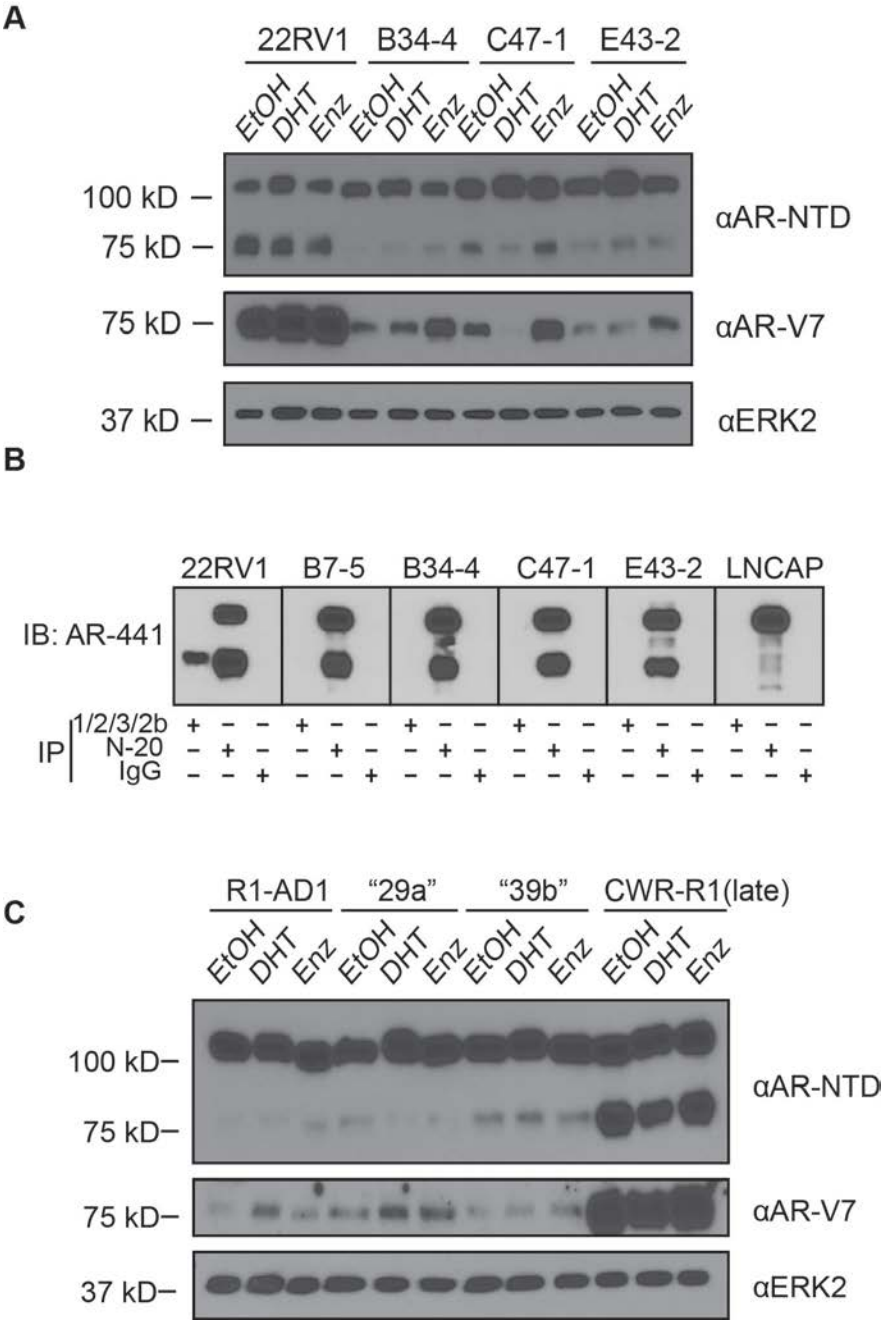


Figure 3-3 (*A*) AR-NTD and AR-V7 Western blot of 22Rv1 and gene-corrected subclones. Cells were cultured for four days in RPMI-1640 media + 10% charcoal stripped serum (CSS) and either 1nM dihydrotestosterone (DHT), 10uM enzalutamide (Enz), or ethanol (EtOH) vehicle. Blots were probed with an anti-AR NTD polyclonal antibody or an anti-AR-V7 mono-clonal antibody. ERK2 functioned as the loading control. (*B*) Immunoprecipitation-western blots of 22RV1, gene corrected 22Rv1 subclones, and LnCaP lysates for AR-V species 1/2/3/2b. (*C*) AR-NTD and AR-V7 western blot of R1-AD1, subclones with a 50kb intron-1 deletion, and CWR-R1 (late passage). As in panel *A*, cells were cultured for four days in RPMI-1640 media + 10% charcoal stripped serum (CSS) and either 1nM dihydrotestosterone (DHT), 10uM enzalutamide (Enz), or ethanol (EtOH) vehicle. Blots were probed with an anti-AR NTD polyclonal antibody or an anti-AR-V7 mono-clonal antibody. ERK2 functioned as the loading control.

Figure 3-4 Androgen dependent growth of *AR* gene modified cell lines

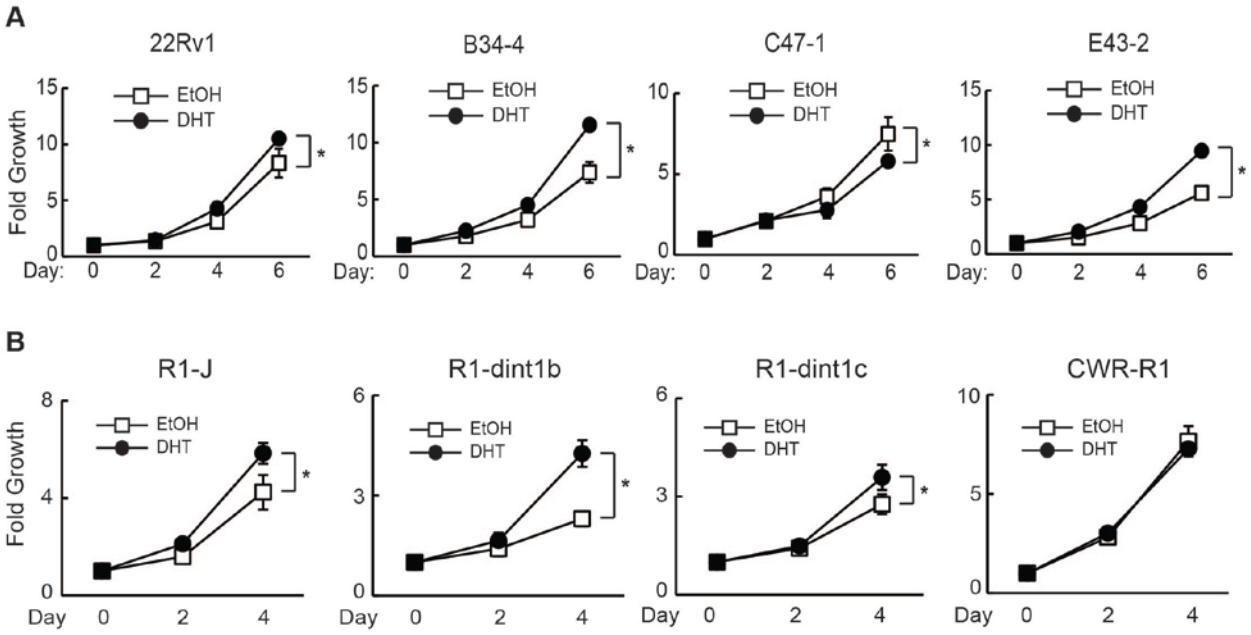


Figure 3-4 (*A*) Six day crystal violet growth curve of 22Rv1 and *AR* gene corrected subclones. Cells were grown in RPMI-1640 + 10% CSS with either 1nM DHT or ethanol vehicle. Error bars represent standard deviation of the mean. These data are derived from a quadruplicate experiment representative of two independent biological replicates (n=4). Asterisks indicate p-values < 0.05. (*B*) Four day crystal violet growth assays of R1-AD1 and gene modified subclones. Cells were grown in RPMI-1640 + 10% CSS with either 1nM DHT or EtOH vehicle. Error bars represent standard deviation of the mean. These data are derived from a sextuplet experiment representative of two independent biological replicates (n=6). Asterisks indicate p-values < 0.05.

Figure 3-5 AR-V7 enhances AR-FL signaling and contributes to androgen-independent growth in a dose dependent manner.

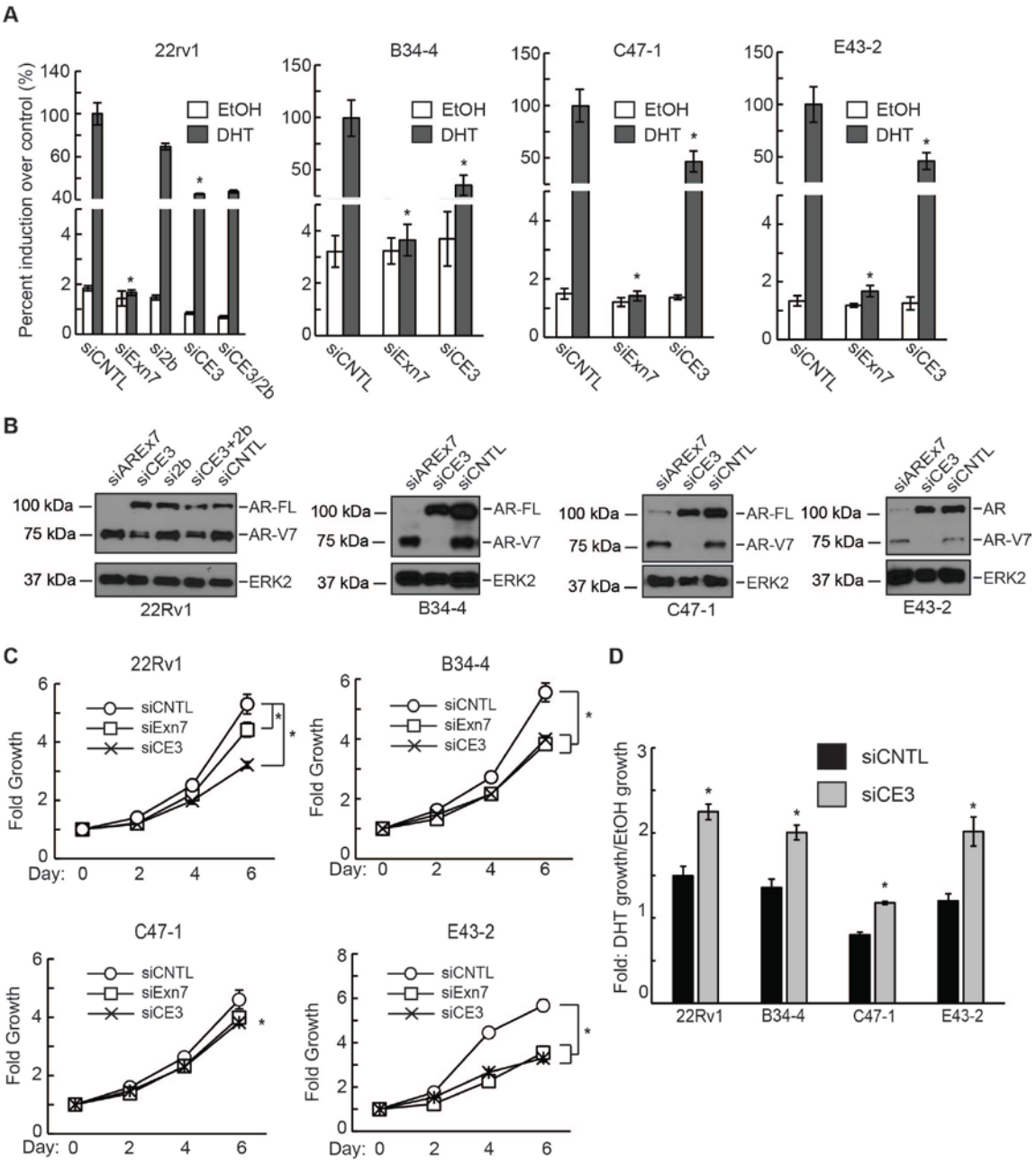


Figure 3-5 AR-V7 enhances AR-FL signaling and contributes to androgen-independent growth. (*A*) AR transcriptional activity in 22Rv1 and modified subclones is indicated by induction of the 4xARE-luciferase reporter in response to 1nM DHT or EtOH vehicle. Luciferase signals were normalized to a co-transfected SV40-Renilla reporter. Cells were transfected with the reporters and siRNAs targeting AR-FL (siExn7), AR-V7 (siCE3), or a non-targeting control (siCNTL). Error bars represent standard deviation. These data are derived from a quadruplicate experiment representative of two independent biological replicates (n=4). Asterisks indicate p-values < 0.05 in comparison to control. (*B*) Western blots confirming knockdown of AR-FL or AR-Vs. (*C*) Crystal violet growth assays of 22Rv1 and subclones under androgen deplete culture conditions with siRNA knockdowns. Error bars represent standard error. These data represent two combined biological replicates (n=8). Asterisks indicate p-values < 0.05. (*D*) The magnitude of androgen induced growth is plotted as the (fold androgenic growth) / (fold castrate growth). Cells were transfected with either siCNTL or siCE3 then cultured for six days in RPMI-1640 + 10% CSS media with either 1nM DHT or EtOH. Cell growth was determined by crystal-violet staining. Error bars represent standard error. These data are two combined biological replicates (n=8). Asterisks indicate p-values < 0.05 when compared to control conditions.

3.6 Tables

Table 3-1 Primers used in study

Genomic PCR			
AR Int2 F1	ctcccaacaagtgatcagtagtcagaaaatgg		
AR Int2 R1	caacagggtatcttattttgcaaaccctaagtc		
AR Int1 F1	gaagtgactgcatatcacgtcatg		
AR Int1 R1	gaattactgacaccaacccaaagc		
AR Int1 R2	caagtccaacacagtttccattacc		

Table 3-2 TALEN targets

TALENs	Target
tAR2	gaacattcctgcctggctgacatgtggactctctgaaattgttat
tAR1-1	ctacttcagtcctttcccacccagctggtttaggaatcaaattccca
tAR1-2	agcatttagctgaattatatattggaagtaagctccttccatgtgg

Table 3-3 TALEN design

RVD position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
tAR2 Left	NN	NI	NI	HD	NI	NG	NG	HD	HD	NG	NN	HD	HD	NG	NN	
tAR2 Right	NI	NG	NI	NI	HD	NI	NI	NG	NG	NG	HD	NI	NN	NI	NN	
tAR1-1 Left	HD	NG	NI	HD	NG	NG	HD	NI	NN	NG	HD	HD	NG	NG	NG	
tAR1-1 Right	NG	NN	NN	NN	NI	NI	NG	NG	NG	NN	NI	NG	NG	HD	HD	NG
tAR1-2 Left	NI	NN	HD	NI	NG	NG	NG	NI	NN	HD	NG	NN	NI	NI	NG	NG
tAR1-2 Right	HD	HD	NI	HD	NI	NG	NN	NN	NI	NI	NN	NN	NI	NN	HD	NG

References

Andersen RJ, Mawji NR, Wang J, et al (2010) Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. *Cancer Cell* **17**(6):535–46

Ateeq B, Tomlins SA, Laxman B, et al (2011) Therapeutic targeting of SPINK1-positive prostate cancer. *Sci Transl Med* **2;3**(72):72ra17.

Attard G, *et al.* (2008) Phase I clinical trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant prostate cancer commonly remains hormone driven. *J Clin Oncol* **26**(28):4563-4571.

Attard G, Richards J, & de Bono JS (2011) New strategies in metastatic prostate cancer: targeting the androgen receptor signaling pathway. *Clin Cancer Res* **17**(7):1649-1657.

Balbas MD, *et al.* (2013) Overcoming mutation-based resistance to antiandrogens with rational drug design. *eLife* **2**:e00499.

Barbieri CE, Baca SC, Lawrence MS, et al (2012) Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature Genetics* **44**(6):685–9

Beltran H, Yelensky R, Frampton GM, *et al* (2012) Targeted Next-generation Sequencing of Advanced Prostate Cancer Identifies Potential Therapeutic Targets and Disease Heterogeneity. *European Urology* pii: S0302-2838(12)01006-8

Berger MF, Lawrence MS, Demichelis F, et al (2011) The genomic complexity of primary human prostate cancer. *Nature* **470**(7333):214–20

Bismar T, Yoshimoto M, Duan Q (2012) Interactions and relationships of PTEN, ERG, SPINK1 and AR in castration-resistant prostate cancer. *Histopathology* **60**(4):645–52

Bogdanove AJ & Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. *Science* **333**(6051):1843-1846.

Bonomi S, *et al* (2013) Oncogenic Alternative Splicing Switches: Role in Cancer Progression and Prospects for Therapy. *International Journal of Cell Biology* doi: 10.1155/2013/962038

Brenner JC, Chinnaiyan AM (2011) Disruptive events in the life of prostate cancer. *Cancer Cell* **19**(3):301–3

Brooke GN, Bevan CL (2009) The role of androgen receptor mutations in prostate cancer progression. *Current Genomics* **10**(1):18–25

Cai C, Balk SP (2011) Intratumoral androgen biosynthesis in prostate cancer pathogenesis and response to therapy. Endocrine-Related Cancer 18(5):R175–82

Cai C, *et al.* (2011) Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. *Cancer Cell* **20**(4):457-471.

Callewaert L, Van Tilborgh N, & Claessens F (2006) Interplay between two hormone-independent activation domains in the androgen receptor. *Cancer Res*. **66**(1):543-553.

Carlson DF, *et al.* (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* **109**(43):17382-17387.
Cermak T, *et al.* (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**(12):e82.
Chan SC, Li Y, Dehm SM (2012) Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal. *The Journal of Biological Chemistry* **287**(23):19736–49
Chang H, Li R, Kuri B, et al. (2013) A gain-of-function mutation in DHT synthesis in castration-resistant prostate cancer. Cell 29;154(5):1074-84.
Chen CD, Welsbie DS, Tran C, et al (2004) Molecular determinants of resistance to antiandrogen therapy. *Nature Medicine* **10**(1):33–9
Chen Y, Clegg NJ, & Scher HI (2009) Anti-androgens and androgen-depleting therapies in prostate cancer: new agents for an established target. *Lancet Oncol* **10**(10):981-991.
Chng KR, Chang CW, Tan SK, et al (2012) A transcriptional repressor co-regulatory network governing androgen response in prostate cancers. *The EMBO Journal* **31**(12):2810–23
Christiaens V, *et al.* (2002) Characterization of the two coactivator-interacting surfaces of the androgen receptor and their relative role in transcriptional control. *J Biol Chem* **277**(51):49230-49237.
Clegg NJ, *et al.* (2012) ARN-509: a novel antiandrogen for prostate cancer treatment. *Cancer Res* **72**(6):1494-1503.
de Bono JS, *et al.* (2011) Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med* **364**(21):1995-2005.
Dehm SM, Regan KM, Schmidt LJ, & Tindall DJ (2007) Selective role of an NH2-terminal WxxLF motif for aberrant androgen receptor activation in androgen depletion independent prostate cancer cells. *Cancer Res* **67**(20):10067-10077.
Dehm SM & Tindall DJ (2007) Androgen receptor structural and functional elements: role and regulation in prostate cancer. *Mol Endocrinol* **21**(12):2855-2863.
Dehm SM, Schmidt LJ, Heemers HV, et al (2008) Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Research* **68**(13):5469–77
Dehm SM, Tindall DJ (2011) Alternatively spliced androgen receptor variants. *Endocrine-Related Cancer* **18**(5):R183–96
DeVore NM & Scott EE (2012) Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. *Nature* **482**(7383):116-119.
Friedlander TW, Roy R, Tomlins S, et al (2012) Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. *Cancer Research* **72**(3):616–25
Gao N, Zhang J, Rao M, et al (2003) The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Molecular Endocrinology* **17**(8):1484–507
Garraway L, Sellers WR (2006) Lineage dependency and lineage-survival oncogenes in human cancer. *Nature Reviews Cancer* **6**(8):593–602

Gottlieb B, Beitel LK, Nadarajah A, et al (2012) The androgen receptor gene mutations database: 2012 update. *Human Mutation* **33**(5):887–94

Grasso CS, Wu Y-M, Robinson DR, et al (2012) The mutational landscape of lethal castration resistant prostate cancer. *Nature* **487**(7406):239-43

Gregory CW, Johnson RT, Jr., Mohler JL, French FS, & Wilson EM (2001) Androgen receptor stabilization in recurrent prostate cancer is associated with hypersensitivity to low androgen. *Cancer Res* **61**(7):2892-2898.

Guo Z, Yang X, Sun F, et al (2009) A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Research* **69**(6):2305–13

Guo C, Chang CC, Wortham M, et al (2012) Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A* **109**(43):17603–8

Haffner MC, Aryee MJ, Toubaji A, et al (2010) Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nature Genetics* **42**(8):668–75

He B, *et al.* (2004) Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* **16**(3):425-438

Hörnberg E, Ylitalo EB, Crnalic S, et al (2011) Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. *PloS One* **6**(4):e19059

Hu R, Dunn T a, Wei S, et al (2009) Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Research* **69**(1):16–22

Hu R, Lu C, Mostaghel E, et al (2012) Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Research* **72**(14):3457–62

Huang S, Gulzar ZG, Salari K, et al (2012) Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. *Oncogene* **31**(37):4164–70

Jernberg E, Thysell E, Ylitalo EB, et al (2013) Characterization of Prostate Cancer Bone Metastases According to Expression Levels of Steroidogenic Enzymes and Androgen Receptor Splice Variants. *Plos One* **8**(11):e77407

Jenster G, van der Korput HA, Trapman J, Brinkmann AO (1995) Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J Biol Chem* **270**(13):7341-7346.

Kumar A, *et al.* (2011) Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A* **108**(41):17087-17092.

Kwon JE, La M, Oh KH, et al (2006) BTB domain-containing speckle-type POZ protein (SPOP) serves as an adaptor of Daxx for ubiquitination by Cul3-based ubiquitin ligase. *The Journal of Biological Chemistry* **281**(18):12664–72

Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14):1754-1760.

Li C, Ao J, Fu J, et al (2011) Tumor-suppressor role for the SPOP ubiquitin ligase in signal-dependent proteolysis of the oncogenic co-activator SRC-3/AIB1. *Oncogene* **30**(42):4350–64

Li Y, Alsagabi M, Fan D, et al (2011) Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. *Cancer Research* **71**(6):2108–17

Li Y, Hwang TH, Oseth L, et al (2012) AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. *Oncogene* **8**;31(45):4759-67

Li Y, et al. (2013) Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res* **73**(2):483-489.

Lin D, Fang H, Ma A, et al (2004) Negative Modulation of Androgen Receptor Transcriptional Activity by Daxx. *Mol Cell Biol* **24**(24):10529-41.

Lin C, Yang L, Tanasa B, et al (2009) Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell 139(6):1069–83

Liu LL, *et al.* (2013) Mechanisms of the androgen receptor splicing in prostate cancer cells. *Oncogene* doi: 10.1038/onc.2013.284

Lupien M, Eeckhoute J, Meyer C, et al (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132(**6):958–70

Lupien M, Brown M (2009) Cistromics of hormone-dependent cancer. *Endocrine-Related Cancer* **16(**2):381–9

Mani R-S, Tomlins S, Callahan K, et al (2009) Induced chromosomal proximity and gene fusions in prostate cancer. *Science* **326**(5957):1230

Mehra R, Han B, Tomlins S, et al (2007) Heterogeneity of TMPRSS2 gene rearrangements in multifocal prostate adenocarcinoma: molecular evidence for an independent group of diseases. *Cancer Research* **67**(17):7991–5

Miller JC, *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**(2):143-148.

Miyamoto DT, *et al.* (2012) Androgen receptor signaling in circulating tumor cells as a marker of hormonally responsive prostate cancer. *Cancer discovery* **2**(11):995-1003.

Montgomery RB, et al (2008) Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. *Cancer Res* **68**(11):4447-4454.

Mostaghel E, Marck BT, Plymate SR, et al (2011) Resistance to CYP17A1 inhibition with abiraterone in castration-resistant prostate cancer: induction of steroidogenesis and androgen receptor splice variants. *Clinical Cancer Research* 17(18):5913–25

Nyquist MD, Dhem SM (2013) Interplay between genomic alterations and androgen receptor signaling during prostate cancer development and progression. *Hormones and Cance*r **4**(2):61-9

Nyquist MD, Li Y, Hwang TH, et al (2013) TALEN-engineered AR gene rearrangements reveal endocrine uncoupling of androgen receptor in prostate cancer. *Proc Natl Acad Sci U S A* **110**(43):17492-7

Pienta KJ, Bradley D (2006) Mechanisms underlying the development of androgen-independent prostate cancer. *Clinical Cancer Research* **12**(6):1665–71

Richly H, Aloia L, Di Croce L (2011) Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death & Disease* **2**(9):e204

Rubin M, Maher C, Chinnaiyan AM (2011) Common gene rearrangements in prostate cancer. Journal of Clinical *Oncology* **29**(27):3659–68

Ruiz C, Lenkiewicz E, Evers L, et al (2011) Advancing a clinically relevant perspective of the clonal nature of cancer. *Proc Natl Acad Sci U S A* **19**;108(29)

Ryan CJ & Tindall DJ (2011) Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. *J Clin Oncol* **29**(27):3651-3658.

Ryan CJ, *et al.* (2012) Abiraterone in Metastatic Prostate Cancer without Previous Chemotherapy. *N Engl J Med* **10**;368(2):138-48

Saraon P, Jarvi K, Diamandis EP (2011) Molecular alterations during progression of prostate cancer to androgen independence. *Clinical Chemistry* **57**(10):1366–75

Scher HI, Fizazi K, Saad F, et al (2012) Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med* **367**(13):1187–97

Shafi AA, Yen AE, Weigel NL (2013) Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacology and Therapeutics* **140**(3):223-38

Shin S, Kim T-D, Jin F, et al (2009) Induction of prostatic intraepithelial neoplasia and modulation of androgen receptor by ETS variant 1/ETS-related protein 81. *Cancer Research* **69**(20):8102–10

Siegel R, Naishadham D, Jemal A (2012) Cancer Statistics, 2012. *CA Cancer J Clin.* **62**(5):283-98

Siegel R, Naishadham D, & Jemal A (2013) Cancer statistics, 2013. *CA Cancer J Clin* **63**(1):11-30.

Steinkamp MP, O'Mahony O, Brogley M, et al (2009) Treatment-dependent androgen receptor mutations in prostate cancer exploit multiple mechanisms to evade therapy. *Cancer Research* **69**(10):4434–42

Subramanian A, *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**(43):15545-15550.

Sun S, Sprenger CCT, Vessella RL, et al (2010) Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* **120**(8):2715-30.

Taylor BS, Schultz N, Hieronymus H, et al (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**(1):11–22

Tepper CG, Boucher DL, Ryan PE, et al (2002) Characterization of a Novel Androgen Receptor Mutation in a Relapsed CWR22. *Cancer Research* **62**(22):6606-14.

Titus MA, Schell MJ, Lih FB, Tomer KB, & Mohler JL (2005) Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. *Clin Cancer Res* **11**(13):4653-4657.

Tomlins S, Laxman B, Dhanasekaran SM, et al (2007) Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**(7153):595–9

Tomlins S, Rhodes D, Yu J, et al (2008) The Role of SPINK1 in ETS Rearrangement-Negative Prostate Cancers. *Cancer Cell* **13**(6):519–528

Tran C, *et al.* (2009) Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**(5928):787-790.

Wang Q, Carroll JS, Brown M (2005) Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell* **19**(5):631–42

Wang Q, Li W, Liu XS, Carroll JS, et al (2007) A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Molecular Cell* **27**(3):380–92

Watson P, Chen YF, Balbas MD, et al (2010) Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc Natl Acad Sci U S A* **107**(39):16759–65

Weichert W, Röske, Gekeler V, et al (2008) Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. *British Journal of Cancer* **98**(3):604–10

Yu J, Yu J, Mani R-S, et al (2010) An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**(5):443–54

Yu Z, Chen S, Sowalsky A, et al (2014) Rapid induction of Androgen Receptor Splice Variants by Androgen Deprivation in Prostate Cancer. *Clin Cancer Res* DOI: 10.1158/1078-0432.CCR-13-1863

Zhang X, Morrissey C, Sun S, et al (2011) Androgen receptor variants occur frequently in castration resistant prostate cancer metastases. *PloS One* **6**(11):e27970

Zhao H, *et al.* (2012) Transcript levels of androgen receptor variant AR-V1 or AR-V7 do not predict recurrence in patients with prostate cancer at indeterminate risk for progression. *J Urol* **188**(6):2158-2164.

Zhou ZX, Sar M, Simental JA, et al (1994) A Ligand-dependent Bipartite Nuclear Targeting Signal in the Human Androgen Receptor. *The Journal of Biological Chemistry* **269**(18):13115-23.